

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 16, 2003, 22:51:47 ; Search time 459 Seconds  
(without alignments)  
3158.166 Million cell updates/sec

Title: US-09-712-338-2\_COPY\_19\_555  
Perfect score: 2887  
Sequence: 1 LPGSTPASVGRRLPKNPCTG.....HTQSSVFLPTATSMSSVGM 537

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DRV=xlp  
-Q=/cgp2\_1/USPTO.spool\_p/US09712338/runat\_16092003\_144913\_19119/app\_query.fasta\_1.711  
-DB=N\_Geneseq\_19Jun03 -QFMT=fastp -SUFFIX=p2n.rng -MATRIX=blom62 -TRANS=human40.cdi  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=20 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09712338 -CGEN 1.1.490 -runat\_16092003\_144913\_19119 -NCPH=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_19Jun03:\*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*
- 25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	2883	99.9	1662	19	AAV28620	A. oryzae ATCC2038
2	2338	81.0	1656	24	ABZ78288	A. niger serine ca
3	1983	68.7	3150	24	ABZ78231	A. niger serine ca
4	980.5	34.0	1872	24	ABZ78289	A. niger serine ca
5	977	33.8	3221	24	ABZ78232	A. niger serine ca
6	941	32.6	1665	24	ABZ78283	A. niger serine ca
7	916.5	31.7	1581	24	ABZ78241	A. niger serine ca
8	894	31.0	2940	24	ABZ78184	A. niger serine ca
9	887.5	30.7	3080	24	ABZ78226	A. niger serine ca
10	704.5	24.4	1007	25	ABZ51997	Aspergillus oryzae
11	636	22.0	2027	24	ABZ76315	S. cerevisiae BAX-
12	630	21.8	1611	24	ABZ78243	A. niger carboxype
13	620	21.5	2002	17	AAT28284	A. niger SFAG 2 ca
14	580	20.1	2068	17	AAT28283	A. niger Bo-1 geno
15	580	20.1	2660	24	ABZ78186	A. niger carboxype
16	568.5	19.7	2503	15	AAO55347	Sequence of gene X
17	541.5	18.8	1653	24	ABO76547	C. albicans BAX-as
18	474	16.4	1551	24	ABZ12878	Aspergillus thalia
19	460.5	16.0	678	21	AAF12116	Aspergillus oryzae
20	414	14.3	1446	24	ABZ78269	A. niger carboxype
21	383	13.3	1510	21	AAC42522	Arabidopsis thalia
22	372	12.9	626	21	AAF12522	Aspergillus oryzae
23	365	12.6	1368	24	ABZ78267	A. niger carboxype
24	361.5	12.5	1936	21	AAC40187	Arabidopsis thalia
25	361	12.5	1551	19	AAV64076	Human serine carbo
26	359	12.4	1686	22	AAF93808	Human cDNA encodin
27	359	12.4	1428	22	AAF94477	Human hydrophobic
28	359	12.4	1633	22	AAF94487	Human hydrophobic
29	359	12.4	1695	21	AAA47444	Human RANGO 176 co
30	359	12.4	1963	22	AAH99751	Human protein enco
31	359	12.4	2076	22	AAH52243	Protein PRO223 CDN
32	359	12.4	2076	22	AAH61129	Human DNA encoding
33	359	12.4	2076	22	AAH72401	Human PRO223 cDNA.
34	359	12.4	2076	25	ACA54942	Novel human secret
35	359	12.4	2076	25	ACA57887	Human PRO223 cDNA.
36	359	12.4	2076	25	ACA58427	cDNA encoding huma
37	359	12.4	2076	25	ACA60134	Human cDNA for sec
38	359	12.4	2076	25	ABX98357	Human cDNA encodin
39	359	12.4	2076	25	ABX98859	Novel human secret
40	359	12.4	2076	25	ACA05472	cDNA encoding huma
41	359	12.4	2076	25	ACA05904	Human secreted/tra
42	359	12.4	2076	25	ABX96151	Human secreted/tra
43	359	12.4	2076	25	ABX97948	Human PRO polynucl
44	359	12.4	2076	25	ABX78732	Human PRO polynucl
45	359	12.4	2076	25	ABX75745	Human cDNA encodin

ALIGNMENTS

RESULT 1  
AAV28620  
ID AAV28620 standard; DNA; 1662 BP.  
XX AAV28620;  
XX  
XX  
XX  
DT 27-AUG-1998 (first entry)  
XX  
DE A. oryzae ATCC20386 carboxypeptidase I DNA.  
XX  
KW Carboxypeptidase I; flavour improving agent; hydrolysate; proteinaceous;  
XX food industry; ss.  
XX Aspergillus oryzae.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1662



QY 481 ProLeuAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGly 500  
 |||||  
 Db 1495 CCCATCGCTCCTGCAATGTTTAAACCGGACTACTTCGGTGGGATATCCGAGAGGC 1554  
 |||||  
 QY 501 GlnLysLysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGln 520  
 |||||  
 Db 1555 CAGAAGAGATCTGCCAGCTACAGACGAATGGAACGGCTACAGCTACGACATACACAG 1614  
 |||||  
 QY 521 SerSerValProLeuProThrAlaThrSerMetSerValGlyMet 536  
 |||||  
 Db 1615 TCGTCCGTCGCTGCCACGCTACGCTACCGCATGCTCCAGTGTGGTATG 1662  
 |||||

## RESULT 2

ABZ78288

ID ABZ78288 standard; cDNA; 1656 BP.

XX

AC ABZ78288;

XX

DT 24-APR-2003 (first entry)

DE A. niger serine carboxypeptidase cDNA #3.

XX

KW Protease; fungal infection; aspergillosis; food; tanning; detergent;  
 KW protein solubility; viscosity; taste; texture; nutritional value;  
 KW gene; ss.

XX  
 OS Aspergillus niger.

XX Key Location/Qualifiers  
 FH 1..1656  
 CDS /\*tag= a

FT /EC\_number= "3.4.16.6"  
 FT

XX WO200268623-A2.  
 PN

XX 06-SEP-2002.  
 PD

XX 22-FEB-2002; 2002WO-EP01984.  
 PF

XX 23-FEB-2001; 2001EP-0200657.  
 PR

PR 23-FEB-2001; 2001EP-0200658.  
 PR

PR 23-FEB-2001; 2001EP-0200660.  
 PR

PR 26-FEB-2001; 2001EP-0200706.  
 PR

PR 26-FEB-2001; 2001EP-0200707.  
 PR

PR 26-FEB-2001; 2001EP-0200719.  
 PR

PR 28-MAR-2001; 2001EP-0000075.  
 PR

PR 28-MAR-2001; 2001EP-0000078.  
 PR

PR 28-MAR-2001; 2001EP-0000080.  
 PR

PR 28-MAR-2001; 2001EP-0000087.  
 PR

PR 28-MAR-2001; 2001EP-0000088.  
 PR

PR 21-MAY-2001; 2001EP-0000156.  
 PR

PR 21-MAY-2001; 2001EP-0000159.  
 PR

PR 21-MAY-2001; 2001EP-0000160.  
 PR

PR 21-MAY-2001; 2001EP-0000162.  
 PR

PR 21-MAY-2001; 2001EP-0000165.  
 PR

PR 21-MAY-2001; 2001EP-0000166.  
 PR

PR 21-MAY-2001; 2001EP-0000168.  
 PR

PR 21-JUN-2001; 2001EP-0000240.  
 PR

PR 21-JUN-2001; 2001EP-0000242.  
 PR

PR 21-JUN-2001; 2001EP-0000244.  
 PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

16-AUG-2001; 2001EP-0000377.

20-SEP-2001; 2001EP-0000478.

20-SEP-2001; 2001EP-0000483.

22-OCT-2001; 2001EP-0000552.

22-OCT-2001; 2001EP-0000553.

22-OCT-2001; 2001EP-0000554.

22-OCT-2001; 2001EP-0000556.

22-OCT-2001; 2001EP-0000557.

22-OCT-2001; 2001EP-0000558.

15-NOV-2001; 2001EP-0000464.

21-DEC-2001; 2001EP-00005117.

(STAM ) DSM NV.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Edens L, Van Dijk AA,

Krubasik P, Albermann K,

Stock A, Kimpel E;

Klugbauer S, Wagner C,

Fritz A, Von Gustedt W,

Heinrich O;

Maier D, Spreafico F,

Folkers U, Hopper S,

Kemmer W, Tan P;

Stiebler J, Albang R;

WPI; 2002-723203/78.

P-PSDB; ABR38864.

Novel isolated protease polypeptide useful in laboratory, clinical,

pharmaceutical, chemical, diagnostic, personal care and industrial

applications

Claim 1; Page 252-253; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A

polypeptide or polynucleotide of the invention is useful for diagnosing a

fungal infection such as aspergillosis, or as a query sequence to perform

a search against public databases. A polypeptide of the invention is

useful in a selected number of industrial or pharmaceutical processes, in

laboratory or clinical processes, in food industry (baking, brewing, in

cheese manufacturing, meat tenderising), in tanning industry and in the

manufacture of biological detergents. A polypeptide may also be useful

for improving protein solubility, extraction yields, viscosity or taste,

texture, nutritional value, minimising of antigenicity or

anti-nutritional factors, colour or functionality as well as processing

aspects like filterability of the proteinaceous raw material. The

sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the

A. niger proteases of the invention.

SQ Sequence 1656 BP; 411 A; 426 C; 409 G; 410 T; 0 other;

Alignment Scores:

Pred. No.: 4.5e-222 Length: 1656

Score: 2338.00 Matches: 419

Percent Similarity: 88.87% Conservative: 52

Best Local Similarity: 79.06% Mismatches: 59

Query Match: 80.98% Indels: 0

DB: 24 Gaps: 0

US-09-712-338-2\_COPY\_19\_555 (1-537) x ABZ78288 (1-1656)

QY 1 LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly 20

Db 55 ATGCCCGAGATGAATGGTCAATCAAGAGGAGGAGTACCAAGAGGCTCCACTGCC 114

QY 21 ValLysThrLeuThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGlu 40

Db 115 GTCAATCGATAAAACCCCAACATGTCACTATCAGGTATATAAGAACGAGAACGAA 174

QY 41 GlyValCysGluThrThrProGlyValLysSerTyrSerGlyThrValAspThrSerPro 60

Db 175 GSAATTGTGAGACACACCTGGGTCAAATCAATCTCCGATATGCGATCTTCGCCCA 234

QY 61 GluSerHisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAlaProIle 80

Db 235 GAGTCGCACTACTTCTTGGTTTTCGAGTCAGCGCGTACCCCGAAGATGATCAGTG 294

QY 81 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuGluLeu 100

Db	295	ACTCTGTGGCTGAATGGTGGCCCTCGAAGCGATTCTCTGATGGGCTTTTGAAGAGTTG	354
QY	101	GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerThrPheAsnGlu	120
Db	355	GGPCCGTGTACATCACACAGAGTAGAATCAATCAATCAGTACTCTCGACAGAG	414
QY	121	ValSerAsnLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr	140
Db	415	GTACCAAAATCTTTCTTGTCTCAGCCCTCGGTGGGGTCTCTTACAGTGAAC	474
QY	141	ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal	160
Db	475	GAGCGGGTTCCTTGATTCATTTACTGGAGCCGTGAGAACGCCCTCTTGTGGAGTT	534
QY	161	GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGlu	180
Db	535	CAGGGTGCATACCCAGTATATGATGCCACTATATCATCGACACGACCGAATCGCTGCACGC	594
QY	181	AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal	200
Db	595	GCAACCTGGAGGGTCTCAGGGCTTCCTCAGTGGCTCTCGCAGCTAGATTCGGAAGTC	654
QY	201	GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhe	220
Db	655	AAGTCCAAAGAGTTCAACTGTGGACAGAGATTACGGAGGACACTATGGACCAAGCTTC	714
QY	221	PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln	240
Db	715	TTCAATCATTTCTACGAGCAAAATTCGAAGATCTAGCGGGGAAGTCAATGGGGTCCAA	774
QY	241	LeuAsnPheAsnSerLeuGlyIleLeuAsnGlyIleIleAspGluAlaIleGlnAlaPro	260
Db	775	CTGAATTTAACTCCCTCGGGATTTATCAACGGCATCATGTCCGGCGGATTCAGGCACAC	834
QY	261	TyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrVal	280
Db	835	TACTACGCAGACTTTGGCGTTAATAATACATATGAATCAAAAGCTGTCAATGACACAGTG	894
QY	281	TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr	300
Db	895	TACAACATATATGAAGTTCGCCAACACGATGCAATGGATGCCAGATCAGGTGCTTCG	954
QY	301	CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn	320
Db	955	TGTAATATGACCAATAGAACTCGCTTCTTGATATGCTATATGTACAGAACAGCCCAAT	1014
QY	321	MetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAsp	340
Db	1015	ATGTGCGAGGACAAATGTGGAAGGGCTTACTACCAAGTTTGGCGCGCTGTGTATGAT	1074
QY	341	IleArgHisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys	360
Db	1075	ATTCGGCACCCCTACAATGACCCGACCCCGCGTCTTACTTGTGTGACTACCTCAAGAA	1134
QY	361	AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal	380
Db	1135	GACTCAGTCATGATGCTATCGCGTGGACATTAATACACCACTCCAGCGCGCAAGTA	1194
QY	381	TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu	400
Db	1195	TATTATGCAATCCAGACACCGCGGACTTGTATGCGCGAATTTTATTGAGACCTCGAA	1254
QY	401	GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys	420
Db	1255	GAGATCCCAACTCCCCGTACGGTGCTGTGATCTACGGGGATGCCGACATATCTGT	1314
QY	421	AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaGlnPhe	440
Db	1315	AACTGGTTCGGCGGTGAGCCCATCTCAGTCGAGTTAACTACCCCATCGAGTCACTTC	1374
QY	441	ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu	460
Db	1375	CGTGACGGGGATACACCCCATGACAGTAGATGGGGTCAATACAGGTGAGACTCGCAG	1434
QY	461	TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln	480
Db	1435	TATGGCAACTTTTCGTTACCCCGGTATATCAGCGCTGATCAGCGAGGATCCATATCAA	1494
QY	481	ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTyrAspIleAlaGluGly	500
Db	1495	CCGATCCAGCGTTCGACCTGTTCAACCGTACTTTATTTGATGGGATATTGCACCGGTT	1554
QY	501	GlnLysLysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGln	520
Db	1555	ACAACTCAGATTTGGCCCGATATATAGCACCAACGAGCATCGCAGGCTACACACAGGAG	1614
QY	521	SerSerValProLeuProThrAlaThrSer	530
Db	1615	TCGTTCTGTCACACTGTCCACGGCTCGAGT	1644
RESULT 3			
Db	ABZ78231	standard; DNA; 3150 BP.	
XX	AC	ABZ78231;	
XX	24-APR-2003	(first entry)	
XX	A. niger serine carboxypeptidase gene #3.		
XX	Protease; fungal infection; aspergilliosis; food; tanning; detergent;		
XX	Protein solubility; viscosity; taste; texture; nutritional value;		
XX	gene; ds.		
XX	Aspergillus niger.		
XX	WO200268623-A2.		
XX	06-SEP-2002.		
XX	22-FEB-2002; 2002WO-EP01984.		
XX	23-FEB-2001; 2001EP-0200657.		
XX	23-FEB-2001; 2001EP-0200660.		
XX	26-FEB-2001; 2001EP-0200706.		
XX	26-FEB-2001; 2001EP-0200707.		
XX	26-FEB-2001; 2001EP-0200708.		
XX	26-FEB-2001; 2001EP-0200719.		
XX	28-MAR-2001; 2001EP-0000075.		
XX	28-MAR-2001; 2001EP-0000078.		
XX	28-MAR-2001; 2001EP-0000080.		
XX	28-MAR-2001; 2001EP-0000087.		
XX	28-MAR-2001; 2001EP-0000088.		
XX	21-MAY-2001; 2001EP-0000156.		
XX	21-MAY-2001; 2001EP-0000159.		
XX	21-MAY-2001; 2001EP-0000160.		
XX	21-MAY-2001; 2001EP-0000162.		
XX	21-MAY-2001; 2001EP-0000165.		
XX	21-MAY-2001; 2001EP-0000166.		
XX	21-MAY-2001; 2001EP-0000168.		
XX	21-JUN-2001; 2001EP-0000240.		
XX	21-JUN-2001; 2001EP-0000242.		
XX	21-JUN-2001; 2001EP-0000244.		
XX	21-JUN-2001; 2001EP-0000246.		
XX	12-JUL-2001; 2001EP-0000280.		
XX	12-JUL-2001; 2001EP-0000285.		
XX	30-JUL-2001; 2001EP-0000323.		
XX	30-JUL-2001; 2001EP-0000327.		
XX	02-AUG-2001; 2001EP-0000341.		
XX	02-AUG-2001; 2001EP-0000342.		
XX	02-AUG-2001; 2001EP-0000343.		
XX	02-AUG-2001; 2001EP-0000344.		
XX	09-AUG-2001; 2001EP-0000357.		
XX	16-AUG-2001; 2001EP-0000374.		
XX	16-AUG-2001; 2001EP-0000377.		





QY 326 ----- 326  
 Db 2005 CGGTGAGTGTCTACTGTTCTCTGCGGGTCAATGAUCAAGGACTTTGCTAAGCTG 2064  
 QY 327 ----- GluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspTleArgHis 343  
 Db 2065 TCATGTACAGAAGGCCCTACTACACAGTTTGGCGCGTGTGTATGATATCGGCAC 2124  
 QY 344 ProTyrAsp----- 346  
 Db 2125 CCTACAA-TGTAACTGGCAAGATAAAGCATTTCTCCGAACAGGACACTGCTCAT 2183  
 QY 347 ----- AspProThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSer 362  
 Db 2184 ATGTCAACTAGACCCCGCGCTCTACTTGTGACTACTCTCAAGAAAGACTCA 2243  
 QY 363 ValMetAspAlaIleGlyValAsnLysPheLeuTyrThrGlnSerAsnAsnValTyrTyr 382  
 Db 2244 GTCATGGATGCTATGCGCGTGGACATTAACACCGAGTCCAGCGCGGAAGTATATAT 2303  
 QY 383 AlaPheGlnGlnThrGlyAspPheValTyrProAsnLysPheLeuGluGluLeu 402  
 Db 2304 GCATTCACACAGACCGCGCACTTTGTATGCGCAATTCATGAGGACCTCGAAGAGATC 2363  
 QY 403 LeuAlaLeuProValArgValSerLeuLeuTyrGlyAspAlaAspTyrIleCysAsnTrp 422  
 Db 2364 CTCCAACTCCCGTACCGGTGCTGTGATCTAGCGCGATGCCGACTATATCTGTAAGTGG 2423  
 QY 423 PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer 442  
 Db 2424 TTGCGGGTCAAGCCATCTACATCGCAGTTAACTAGCCCATGCTAGTCCGTCGA 2483  
 QY 443 AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 462  
 Db 2484 GCGGGATACACCCATGAGATGAGTGGTGGATACGCTGAGACTCCGAGTAGTGGC 2543  
 QY 463 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProile 482  
 Db 2544 AACTTTCGTTCCCGCGTATATCATCGCTGGCAGGATTCATATCTATCAACCGATC 2603  
 QY 483 AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLys 502  
 Db 2604 GCACGCTGCGAGTGTTCACCGTACTTTATTTGGATGGATATTCGACGCGGTACAACT 2663  
 QY 503 LysIleThrProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer 522  
 Db 2664 CAGATTGGCCGGAATATAGCACCAACGAGGACATCGCAGGTACACACAGGATCGTTC 2723  
 QY 523 ValProLeuProThrAlaThrSer 530  
 Db 2724 GTGCCACTGTCCAGCGGTGAGT 2747

RESULT 4

ABZ78289  
 ID ABZ78289 standard; cDNA; 1872 BP.

XX XX  
 AC ABZ78289;

DT 24-APR-2003 (first entry)

DE A. niger serine carboxypeptidase cDNA #4.

KW protease; fungal infection; aspergillosis; food; tanning; detergent;  
 KW protein solubility; viscosity; taste; texture; nutritional value;  
 KW gene; ss.

XX Aspergillus niger.

XX Key Location/Qualifiers  
 FH 1..1872  
 CDS /\*tag= a  
 FT /EC\_number= "3.4.16.6"  
 FT  
 XX

PN  
 XX  
 PD  
 XX  
 PF  
 XX  
 PR 23-FEB-2001; 2001EP-0200657.  
 PR 23-FEB-2001; 2001EP-0200658.  
 PR 23-FEB-2001; 2001EP-0200660.  
 PR 26-FEB-2001; 2001EP-0200706.  
 PR 26-FEB-2001; 2001EP-0200707.  
 PR 26-FEB-2001; 2001EP-0200708.  
 PR 26-FEB-2001; 2001EP-0200719.  
 PR 28-MAR-2001; 2001EP-0000075.  
 PR 28-MAR-2001; 2001EP-0000080.  
 PR 28-MAR-2001; 2001EP-0000087.  
 PR 28-MAR-2001; 2001EP-0000088.  
 PR 21-MAY-2001; 2001EP-0000156.  
 PR 21-MAY-2001; 2001EP-0000160.  
 PR 21-MAY-2001; 2001EP-0000162.  
 PR 21-MAY-2001; 2001EP-0000165.  
 PR 21-MAY-2001; 2001EP-0000166.  
 PR 21-MAY-2001; 2001EP-0000168.  
 PR 21-JUN-2001; 2001EP-0000240.  
 PR 21-JUN-2001; 2001EP-0000242.  
 PR 21-JUN-2001; 2001EP-0000244.  
 PR 21-JUN-2001; 2001EP-0000246.  
 PR 12-JUL-2001; 2001EP-0000280.  
 PR 12-JUL-2001; 2001EP-0000285.  
 PR 30-JUL-2001; 2001EP-0000323.  
 PR 30-JUL-2001; 2001EP-0000327.  
 PR 02-AUG-2001; 2001EP-0000341.  
 PR 02-AUG-2001; 2001EP-0000342.  
 PR 02-AUG-2001; 2001EP-0000343.  
 PR 02-AUG-2001; 2001EP-0000344.  
 PR 09-AUG-2001; 2001EP-0000357.  
 PR 16-AUG-2001; 2001EP-0000374.  
 PR 16-AUG-2001; 2001EP-0000377.  
 PR 20-SEP-2001; 2001EP-0000478.  
 PR 20-SEP-2001; 2001EP-0000552.  
 PR 22-OCT-2001; 2001EP-0000553.  
 PR 22-OCT-2001; 2001EP-0000554.  
 PR 22-OCT-2001; 2001EP-0000556.  
 PR 22-OCT-2001; 2001EP-0000557.  
 PR 15-NOV-2001; 2001EP-0000558.  
 PR 21-DEC-2001; 2001EP-0004464.  
 PR 21-DEC-2001; 2001EP-0005117.

(STAM ) DSM NV.

Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;  
 Klugbauer S, Wagner C, Fritz A, Von Gustadt W, Heinrich O;  
 Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;  
 Stiebler J, Albang R;

WPI; 2002-723203/78.  
 P-PSDB; ABR38865.

Novel isolated protease polypeptide useful in laboratory, clinical,  
 pharmaceutical, chemical, diagnostic, personal care and industrial  
 applications -

Claim 1; Page 253-254; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A  
 polypeptide or polynucleotide of the invention is useful for diagnosing a  
 fungal infection such as aspergillosis, or as a query sequence to perform  
 a search against public databases. A polypeptide of the invention is  
 useful in a selected number of industrial or pharmaceutical processes, in  
 laboratory or clinical processes, in food industry (baking, brewing,

CC cheese manufacturing, meat tenderising), in tanning industry and in the  
 CC manufacture of biological detergents. A polypeptide may also be useful  
 CC for improving protein solubility, extraction yields, viscosity or taste,  
 CC texture, nutritional value, minimising of antigenicity or  
 CC anti-nutritional factors, colour or functionality as well as processing  
 CC aspects like filterability of the proteinaceous raw material. The  
 CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the  
 CC A. niger proteases of the invention.

XX  
 SQ Sequence 1872 BP; 402 A; 434 C; 570 G; 466 T; 0 other;

#### Alignment Scores:

Pred. No.:	3,22e-87	Length:	1872
Percent:	980.50	Matches:	216
Best Local Similarity:	56.23%	Conservative:	91
Query Match:	39.56%	Mismatches:	178
DB:	33.96%	Indels:	61
	24	Gaps:	19

US-09-712-338-2\_COPY\_19\_555 (1-537) x ABZ78289 (1-1872)

QY	26	ThrAlaAsnAsnValThrIleArgTyrLysGlu-----ProGlyAlaGlu-----	40
Db	94	ACGCTGAGGATCTCACGGTATTCAATCGGAGATATTCCCTGGTGGGAGGATCTCCPAT	153
QY	41	-----GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrVal	56
Db	154	AAGCAACCCCTCGGCATCTGCACCAACCCACCCCTCCACCCCGGCTACTCGGGTACATC	213
QY	57	AspThrSerProGluSer-----	62
Db	214	CACCTCCCGCCACACCCCTTACCATTCTCTCCATTCAGGAATCAGCATCTCGCAACCA	273
QY	63	-----HisThrPhePheTrpPheGluAlaArg-----HisAsnProGluThr	77
Db	274	TACCTATCAATACCTTTCTGGTACTTCTCTCCCGCCATCCCAACAATGATACA	333
QY	78	AlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPhe	97
Db	334	TCGCCATCACCATCTGGATGAACGGCGGCGCGCGGATCTCCATGATGGGCTATT	393
QY	98	GluGluLeuGlyProCysHisValAsnSerThrPheAspAsp-----TyrIleAsnProHis	116
Db	394	CAGAGAACGGGGCATGACTGTGAATACGGACTCGAATCCACGGCTATAATCCCTGG	453
QY	117	SerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer	136
Db	454	TCGTGGAATGAGTACGTCGATATCTGATATGAGCAGCGCGGTGCGACGGGATTTAGT	513
QY	137	TyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSer	156
Db	514	TATCATGTGTGTGAGGAATGGAGGTATAGAT-----TTGAATGAGACG	555
QY	157	PheAlaGlyValGlnGlyArgTyrProThrIleAsp-----AlaThrLeuIleAspThr	174
Db	556	TTT-----TTGTGGGGAGGTTCGGGATCAGATCAGATGTGCATGGGAGGTGAATGGGAG	609
QY	175	ThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGly-----	192
Db	610	GTTAAT--GGGGAGGGCGCTTGGGTGGCTTCAGAGTTGGTGGGTGGTGAATCTCT	666
QY	193	-----LeuProSerLeuAspSerArgValGlnSerLysAsp---PheSerLeuTrpThr	209
Db	667	GAATATGTTCTCTCTGTCAGGGAATGGTGGTGGTGCATCAGAGGGTGCATATGGACG	726
QY	210	GluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGluGlnAsnGlu	229
Db	727	GAGTCATATGGGGAGCGGTATGACCGGATACACGGCGCTCTTTCAGGAGATGAATGAG	786
QY	230	ArgIleAlaAsnGlySerValAsn---GlyValGlnLeuAsnPheAsnSerLeuGlyIle	248
Db	787	AGGATTCAGATGGGGAGTGAACCGGGAAGAGATCCATTTGATACGCTGGGCATT	846

QY	249	IleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAlaValAsn	268
Db	847	ATCAATGGGTGGTGGATTACTCGTCAGAGTCCCTTCGTTCCCTCAGCAGGCTAATAC	906
QY	269	AsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsn	288
Db	907	ANTAGTATGGGATCGAGGGAATCATCCACGCTACACGGGCTATGATATGATGG	966
QY	289	GlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAla	308
Db	967	AGCAAGCCTGGCGGGTGCAGGATATGATCATCGAGTGTCTGGCGATCTGGCGAG	1020
QY	309	LeuAlaAsp---TyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGlu	327
Db	1021	CTCGGAGATCCCTCATCATCTCGGAGGAGGTCTGACTACTGTTCGGGGAGATCAAG	1080
QY	328	GlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyrAsp	347
Db	1081	AGCCTGATACCAATACCTCCGGGGAGGATACIAGACATACGCAATTCACGGCGGAT	1140
QY	348	ProThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIle	367
Db	1141	GCAGCTCTCGTCCCTTACTTCGCTGCTTCTTGAATCCCATCGGCTGCAAAAGCACTT	1200
QY	368	GlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPheGlnThr	387
Db	1201	GGGTCCCGGTGAATATACCATCTGCTCAGAGGAGTGGGACAGTTCGCTCGCAGG	1260
QY	388	GlyAspPheValTrpPro-----AsnPheIleGluAspLeuGluIle	402
Db	1261	GGCGATATAT-----CCGCGAATGATCCCGCGGAATGATCGGGGATATTGGATCTG	1314
QY	403	LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp	422
Db	1315	CTTGACTCCGGTGTCAAGTGGCTATGCTATATGSGGACCGGACTATGCTTGTCCGTGG	1374
QY	423	PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer	442
Db	1375	CGCGCGGGGAGATGTGAGCTCTGCTGGTGGAGTACGAGGATCGCGGAGATTCGCTGCT	1434
QY	443	AlaGlyTyrThrProLeuLysValAsnGlyValGluTyr-----GlyGluThrArgGlu	460
Db	1435	GCTGGTATGCCGAAGTGCACAGCAAGTCA---TCCTACGTGGGGGCTAGTAAGGCAG	1491
QY	461	TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln	480
Db	1492	TATGGGAACITCTGTTACGCGTGTCTTTCAGCGGGCCCATGAGTGCCATTTATCAG	1551
QY	481	ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTyrPheAspIleAlaGluGly	500
Db	1552	CCCGAAACGGGTATGAGATTTTAAATCGCGCTCAGTTTAAATGGGATATTCGACGSGA	1611
QY	501	GlnLysLysIleTrp-----ProSerTyrLysThrAsnGlyThrAlaThrAlaTheHis	518
Db	1612	GGCATTTCTCTGGAGCAGATCAGCTATGGCAGGAGGACCGCTCTGTAACGIGGCAT	1671
QY	519	ThrGlnSerSerValPro	524
Db	1672	ATCAAAACCAAGTGGCG	1689

#### RESULT 5

ABZ78232  
 ID ABZ78232 standard; DNA; 3221 BP.

XX ABZ78232;

XX

DT 24-APR-2003 (first entry)

XX

DE A. niger serine carboxypeptidase gene #4.

XX

KW Protease; fungal infection; aspergillosis; food; tanning; detergent;

KW protein solubility; viscosity; taste; texture; nutritional value;

KW gene; ds.



Db 1529 AAGATCCATTGGATACGCGGGCATATCAATGGGTCTGTGGATTACTCGTCGAGTC 1698  
 QY 260 ProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyLeuLysAlaValAsnGluThr 279  
 Db 1689 CCTTCGTTCCCTGAGCAGCGGTATAACAATACGTATGGATPCGAGGAAICAAATFCGCACG 1748  
 QY 280 ValTyrAsnTyrMetLysPheAlaAsnGluMetProAsnGlyCysGlnAspLeuIleSer 299  
 Db 1749 CTCTACGACCGGCTATGGATAGTTGGAGCAAGCCITGGCGGTTCGAGGGATATGATCATC 1808  
 QY 300 ThrCysLysGlnThr 1868  
 Db 1809 GAGTGTGCGCATGCTGGCGAGCTCGGAGATCCCTCATGTATGCGACAATGAGACGGTA 1868  
 QY 309 LeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGly 328  
 Db 1869 -----AATAGCACTCGAGGAGCGCTCGGACTGCTGCGGGAGATCAAGAGC 1919  
 QY 329 ProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro 348  
 Db 1920 CTGTATACGAATACCTCGCGGCGAGGATACAGACATACGATTCACGCCCGGATGCA 1979  
 QY 349 ThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGly 368  
 Db 1980 GCTCTCGTCTTACTTCGCGGGTCTTCAATCGCCCATGGGTGCAAAAGGCACCTGGG 2039  
 QY 369 ValAsnIleAsnTyrThrGlnSerAsnAspValTyrTyrAlaPheGlnGlnThrGly 388  
 Db 2040 GTCCCGGTGAATATACCATGCTGCTCAGGAGCGTGGGAACAGTTTCGCTCGAGCGGC 2099  
 QY 389 AspPheValTrpPro-----AsnPheIleGluAspLeuGluIleLeu 403  
 Db 2100 GATTAT-----CCGGAATGATCCCGGGAATGATCGGGATATGGATACTGCTT 2153  
 QY 404 AlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPhe 423  
 Db 2154 GACTCGGTGTCAAGTGCTATGGTATGGGATGGGACCGGACTATGCTTGTCTGGTGGGC 2213  
 QY 424 GlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAla 443  
 Db 2214 GCGCGGGAAGATGTCAGCTGCTGTGGAGTACGAGGATCGCGAGAAGTTCGCTGCTGT 2273  
 QY 444 GlyTyrThrProLeuLysValAsnGlyValGluTyr-----GlyGluThrArgGluTyr 461  
 Db 2274 GGGTATGCGGAAGTACGACGAGTCA---TCCTACGTGGGGTCTAGTAGGCGAGTAT 2330  
 QY 462 GlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnPro 481  
 Db 2331 GGGAACTTCTGTTCCCGGCTCTTTCAGGCGGCCATGAGTGCCCATTTTATCAGGCC 2390  
 QY 482 IleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTyrAspIleAlaGluGln 501  
 Db 2391 GAAACGGGTATGAGATTTTAATCGCGTCAGTTTATGGATATTCGCGAGGAGGC 2450  
 QY 502 LysLysIleTrp-----ProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThr 519  
 Db 2451 ATTTCTCTGGAGCAAGTACAGAGTATGGGAGCGGAGGACCGTCTGTCACGTTGGATATC 2510  
 QY 520 GlnSerSerValPro 524  
 Db 2511 AAAAAAGAGTGGC 2525  
 RESULT 6  
 ABZ78283  
 ID ABZ78283 standard; cDNA; 1665 bp.  
 XX  
 AC ABZ78283;  
 XX  
 DT 24-APR-2003 (first entry)  
 XX  
 DE A. niger serine carboxypeptidase cDNA #2.  
 XX

KW Protease; fungal infection; aspergillois; food; tanning; detergent;  
 KW protein solubility; viscosity; taste; texture; nutritional value;  
 XX gene; ss.  
 OS Aspergillus niger.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1665  
 FT /\*tag= a  
 FT /EC\_number= "3.4.16.6"  
 XX  
 PN WO200268623-A2.  
 XX  
 PD 06-SEP-2002.  
 XX  
 XX 22-FEB-2002; 2002WO-EP01984.  
 PR 23-FEB-2001; 2001EP-0200657.  
 PR 23-FEB-2001; 2001EP-0200658.  
 PR 23-FEB-2001; 2001EP-0200660.  
 PR 26-FEB-2001; 2001EP-0200706.  
 PR 26-FEB-2001; 2001EP-0200707.  
 PR 26-FEB-2001; 2001EP-0200708.  
 PR 26-FEB-2001; 2001EP-0200719.  
 PR 28-MAR-2001; 2001EP-0000075.  
 PR 28-MAR-2001; 2001EP-0000078.  
 PR 28-MAR-2001; 2001EP-0000080.  
 PR 28-MAR-2001; 2001EP-0000087.  
 PR 28-MAR-2001; 2001EP-0000088.  
 PR 21-MAY-2001; 2001EP-0000156.  
 PR 21-MAY-2001; 2001EP-0000159.  
 PR 21-MAY-2001; 2001EP-0000160.  
 PR 21-MAY-2001; 2001EP-0000162.  
 PR 21-MAY-2001; 2001EP-0000165.  
 PR 21-MAY-2001; 2001EP-0000166.  
 PR 21-MAY-2001; 2001EP-0000168.  
 PR 21-JUN-2001; 2001EP-0000240.  
 PR 21-JUN-2001; 2001EP-0000242.  
 PR 21-JUN-2001; 2001EP-0000244.  
 PR 21-JUN-2001; 2001EP-0000246.  
 PR 12-JUL-2001; 2001EP-0000280.  
 PR 12-JUL-2001; 2001EP-0000285.  
 PR 30-JUL-2001; 2001EP-0000323.  
 PR 30-JUL-2001; 2001EP-0000327.  
 PR 02-AUG-2001; 2001EP-0000341.  
 PR 02-AUG-2001; 2001EP-0000342.  
 PR 02-AUG-2001; 2001EP-0000343.  
 PR 02-AUG-2001; 2001EP-0000344.  
 PR 09-AUG-2001; 2001EP-0000357.  
 PR 16-AUG-2001; 2001EP-0000374.  
 PR 16-AUG-2001; 2001EP-0000377.  
 PR 20-SEP-2001; 2001EP-0000478.  
 PR 20-SEP-2001; 2001EP-0000483.  
 PR 22-OCT-2001; 2001EP-0000552.  
 PR 22-OCT-2001; 2001EP-0000553.  
 PR 22-OCT-2001; 2001EP-0000554.  
 PR 22-OCT-2001; 2001EP-0000556.  
 PR 22-OCT-2001; 2001EP-0000557.  
 PR 22-OCT-2001; 2001EP-0000558.  
 PR 15-NOV-2001; 2001EP-0000464.  
 PR 21-DEC-2001; 2001EP-0005117.  
 XX (STAM ) DSM NV.  
 PA Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;  
 XX Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;  
 PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;  
 PI Stiebler J, Albarg R;  
 XX WPI; 2002-723203/78.  
 DR P-PSDB; ABR38859.  
 XX Novel isolated protease polypeptide useful in laboratory, clinical,  
 PT

PT pharmaceutical, chemical, diagnostic, personal care and industrial  
PT applications

PS Claim 1; Page 246-247; 394pp; English.

XX The invention relates to a novel isolated protease polypeptide. A  
CC polypeptide or polynucleotide of the invention is useful for diagnosing a  
CC fungal infection such as aspergillosis, or as a query sequence to perform  
CC a search against public databases. A polypeptide of the invention is  
CC useful in a selected number of industrial or pharmaceutical processes, in  
CC laboratory or clinical processes, in food industry (baking, brewing,  
CC cheese manufacture, meat tenderising), in tanning industry and in the  
CC manufacture of biological detergents. A polypeptide may also be useful  
CC for improving protein solubility, extraction yields, viscosity or taste,  
CC texture, nutritional value, minimising of antigenicity or  
CC anti-nutritional factors, colour or functionality as well as processing  
CC aspects like filterability of the proteinaceous raw material. The  
CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the  
CC A. niger proteases of the invention.

XX Sequence 1665 BP; 374 A; 455 C; 424 G; 412 T; 0 other;

# Alignment Scores:

Pred. No.: 2,28e-83 Length: 1665  
Score: 941.00 Matches: 209  
Percent Similarity: 51.75% Conservative: 72  
Best Local Similarity: 38.49% Mismatches: 172  
Query Match: 32.59% Indels: 90  
DB: 24 Gaps: 13

US-09-712-338-2\_COPY\_19\_555 (1-537) x ABZ78283 (1-1665)

QY 13 GlnLeuProLysAsnProThrGlyValLysThrLeuThrAla-----AsnAsnVal 30  
DB 49 CAATTCCTCCCGAGCGGAGCATCACTGCTCAAGTCCAAAGTTGCAAGAAAGTG 108  
QY 31 ThrLeuArgThrGlyGluProGlyAlaGluGlyValCysGluThrThrProGlyValLys 50  
DB 109 ACTATTTCTTCAAGAGCCT-----GGATTTGGAAACTAGCGCGGTGTCGA 159  
QY 51 SerTyrSerGlyTyrValAspThrSerProGluSerHisThrPheThrPhePheGlu 70  
DB 160 TCATTATCGGCTATGATACACCTTCCCGCGCTCAACACAGCTCTTTGGTTTTCGAA 219  
QY 71 AlaArgHisAsnProGluThrAlaProLleThrLeuThrLeuAsnGlyGlyProGlySer 90  
DB 220 GCCCGCAAGATCCAGCAATGCGCTTCCGCTGCTCAAGTGGCGGTGCGGTGCG 279  
QY 91 AspSerLeuThrGlyLeuPheGluLeuGlyProCysHisValAsnSer---ThrPhe 109  
DB 280 TCGTCGCTCATGGGCTCCTTGAAGATTAGTCTTGTTCATTCATCAGACTCCAA 339  
QY 110 AspAspTyrThrLeuAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGln 129  
DB 340 ACCACAGTCTCATCTTGGAGTGGCAATGAAGTCAATCTTCTTATTCCTTGACAG 399  
QY 130 ProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnProValThr 149  
DB 400 CCAACTCAAGTGGCTCTCTCATAGCATGTCACCAACAATGGCATTTCACA----- 450  
QY 150 GlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAla 169  
DB 450 ----- 450  
QY 170 ThrLeuThrAspThrThrAsnLeuAlaGluAlaAlaTrpGluLeuLeuGlnGlyPhe 189  
DB 451 -----GCTAATGGAGTGCATTCGCGGCTCACCTCTATGGCAATTCGCGCAACCTGG 504  
QY 190 LeuSerGlyLeuProSerLeu-----AspSerArgValGlnSerLysAspPheSer 206  
DB 505 TTTTTCGAGTTCACACACTACACCAACCAACGATGATGCTGTC-----AGT 549  
QY 207 LeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGlu 226

DB 550 CTCTGGGTGAAGTTACGGAGGCCATTATGTTCCAGGCATCTTTCGGTTCCTCAACAG 609  
QY 227 GlnAsnGluArgIleAlaAsnGlySerVal---AsnGlyValGln---LeuAsnPheAsn 244  
DB 610 CAGATGCAAAATCGCAGAGGGACTGCAGAAGAGCGGTGCACAGTATTGTCATCTCGAC 669  
QY 245 SerLeuGlyIleLeuAsnGlyLeuLeuAspGluAlaIleGlnAlaProTyrThrProGlu 264  
DB 670 ACGCTGGCATTTGAACGGCTTGTGATGTTGATGTCACAGAGAGGCTTACATTACT 729  
QY 265 PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMet 284  
DB 730 TGGCCA-----TACATAACGTA 747  
QY 285 LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr 304  
DB 748 AGCTCGCCCT-----TCTTCATTC 768  
QY 305 AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp 324  
DB 769 AACTCGGAGGCTTTCGATCAGGCCCTCGCTGCGAGCGCTTTGAAGAACCGCAT 828  
QY 325 -----AsnVal----- 326  
DB 829 TCGGCTTGCTCACTCAGGAAGAAATATCTCTGAAATTTCCGAGCGCTTCACATAGAA 888  
QY 327 -----GluGlyPro-----TyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg 342  
DB 889 TGGGAGATGCGCCATCCTACTACACACCTTCAATCGCGGTGTCAGCATCGCC 948  
QY 343 HisProTyrAspAspProThrProSerTyrTyrAsnLysPheLeuAlaLysAspSer 362  
DB 949 CATCTTAAGAACGACCATTCCTGCCAAGCACATGCTCGATATTTGACGAGGATCC 1008  
QY 363 ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyr 382  
DB 1009 GTCTTCGCTCTTGGGTACCACTCAATTCACATCGCTTCGAGTGGCGGTGCTACA 1068  
QY 383 AlaPheGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIle 402  
DB 1069 CAGTTCATAAAACCTTGTATGCTCCAGCGGCTTCTGGATGCAATTGGCTACCTC 1128  
QY 403 LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp 422  
DB 1129 CTCACAGTGTGTAAAGTACACATGATGTCAGAGATCTGTGATTCGCTGCAATTGG 1188  
QY 423 PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer 442  
DB 1189 GTCGGGCGGAAAAGCCAGCTTGCAGTTCCTGATTCCTGATACCCGATTTGCGGAC 1248  
QY 443 AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlyGly 462  
DB 1249 ACGGATACCTCCACCTCTTACGCCGACGGATCAGCGCATGACCGCGCAGCTGGGC 1308  
QY 463 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGluProIle 482  
DB 1309 AACTACAGCTTCACTCGGCTCTCCAGCCGGGATGAGTCCCTCTCCTACAGCGCTC 1368  
QY 483 AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLys 502  
DB 1369 CGCGGATGATGATCTTCATCGCGGCGCATTCACAAAGATATCCCTACCTGCTCTTG 1428  
QY 503 LysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer 522  
DB 1429 GCTGTGTGATGACGAATTCACGTCGGTTCGACCTAAGGATAGTGGCATATCAAGAATATC 1488  
QY 523 ValProLeu 525  
DB 1489 CCTCCTATT 1497  
RESULT 7  
ABZ78241









QY 469 lYrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAs 489  
 Db 1982 GATGAGATGGACATGAGGTTCCTTCATCAACCTTGCTGCGCTGGAGATGTGA 2041  
 QY 489 nArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLysLysIleTrpProSerTyrLy 509  
 Db 2042 GCGGTCATTTGGCGGCAAGATGTGGGACGGGAAAGATTCCTCGTCGAGTTTACA 2101  
 QY 509 sThrAsnGlyThr----- 513  
 Db 2102 GACGGTGGGACGCCCAAGAGTTTACTACCGGGAGGGCAACAGCAGCATTCAGTGGAGGT 2161  
 QY 514 -----AlaThrAlaThrHis---ThrGlnSerSerValProLeuProThrAlaThrSe 530  
 Db 2162 GTTGGATTCCTGGCGACGTACACACACACACACAGATGCTCCGAACCGGTGACCGGAG 2221  
 QY 530 rMetSerSerValGlyMetAla 537  
 Db 2222 GCTGAAGCGGATGGGACGAGCT 2243

RESULT 9  
 ABZ78226  
 ID ABZ78226 standard; DNA; 3080 BP.  
 XX AC ABZ78226;  
 XX DT 24-APR-2003 (first entry)  
 XX DE A. niger serine carboxypeptidase gene #2.  
 XX KW Protease; fungal infection; aspergillosis; food; tanning; detergent;  
 KW protein solubility; viscosity; taste; texture; nutritional value;  
 KW gene; ds.  
 CS Aspergillus niger.  
 PN WO200268623-A2.  
 XX PD 06-SEP-2002.  
 XX PF 22-FEB-2002; 2002WO-EP01984.  
 XX PR 23-FEB-2001; 2001EP-0200657.  
 PR 23-FEB-2001; 2001EP-0200658.  
 PR 23-FEB-2001; 2001EP-0200660.  
 PR 26-FEB-2001; 2001EP-0200706.  
 PR 26-FEB-2001; 2001EP-0200707.  
 PR 26-FEB-2001; 2001EP-0200708.  
 PR 26-FEB-2001; 2001EP-0200719.  
 PR 28-MAR-2001; 2001EP-0000075.  
 PR 28-MAR-2001; 2001EP-0000078.  
 PR 28-MAR-2001; 2001EP-0000080.  
 PR 28-MAR-2001; 2001EP-0000087.  
 PR 28-MAR-2001; 2001EP-0000088.  
 PR 21-MAY-2001; 2001EP-0000156.  
 PR 21-MAY-2001; 2001EP-0000159.  
 PR 21-MAY-2001; 2001EP-0000160.  
 PR 21-MAY-2001; 2001EP-0000162.  
 PR 21-MAY-2001; 2001EP-0000165.  
 PR 21-MAY-2001; 2001EP-0000166.  
 PR 21-MAY-2001; 2001EP-0000168.  
 PR 21-JUN-2001; 2001EP-0000240.  
 PR 21-JUN-2001; 2001EP-0000242.  
 PR 21-JUN-2001; 2001EP-0000244.  
 PR 21-JUN-2001; 2001EP-0000246.  
 PR 12-JUL-2001; 2001EP-0000280.  
 PR 12-JUL-2001; 2001EP-0000285.  
 PR 30-JUL-2001; 2001EP-0000323.  
 PR 30-JUL-2001; 2001EP-0000327.  
 PR 02-AUG-2001; 2001EP-0000341.  
 PR 02-AUG-2001; 2001EP-0000342.  
 PR 02-AUG-2001; 2001EP-0000343.

PR 02-AUG-2001; 2001EP-0000344.  
 PR 09-AUG-2001; 2001EP-0000357.  
 PR 16-AUG-2001; 2001EP-0000374.  
 PR 16-AUG-2001; 2001EP-0000377.  
 PR 20-SEP-2001; 2001EP-0000478.  
 PR 20-SEP-2001; 2001EP-0000483.  
 PR 22-OCT-2001; 2001EP-0000552.  
 PR 22-OCT-2001; 2001EP-0000553.  
 PR 22-OCT-2001; 2001EP-0000554.  
 PR 22-OCT-2001; 2001EP-0000556.  
 PR 22-OCT-2001; 2001EP-0000557.  
 PR 22-OCT-2001; 2001EP-0000558.  
 PR 13-NOV-2001; 2001EP-0004464.  
 PR 21-DEC-2001; 2001EP-0005117.  
 XX (STAM ) DSM NV.  
 XX PA Edens L, Van Dijk AA, Krubasik P, Albertmann K, Stock A, Kimpel E;  
 PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;  
 PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;  
 PI Stiebler J, Albarg R;  
 XX WPI: 2002-723203/78.  
 DR P-PSDB: ABR38859.  
 DR Novel isolated protease polypeptide useful in laboratory, clinical,  
 PT pharmaceutical, chemical, diagnostic, personal care and industrial  
 PT applications  
 XX Claim 1; Page 178-180; 394pp; English.  
 XX The invention relates to a novel isolated protease polypeptide. A  
 CC polypeptide or polynucleotide of the invention is useful for diagnosing a  
 CC fungal infection such as aspergillosis, or as a query sequence to perform  
 CC a search against public databases. A polypeptide of the invention is  
 CC useful in a selected number of industrial or pharmaceutical processes, in  
 CC laboratory or clinical processes in food industry (baking, brewing,  
 CC cheese manufacturing, meat tenderising), in tanning industry and in the  
 CC manufacture of biological detergents. A polypeptide may also be useful  
 CC for improving protein solubility, extraction yields, viscosity or taste,  
 CC texture, nutritional value, minimising of antigenicity or  
 CC anti-nutritional factors, colour or functionality as well as processing  
 CC aspects like filterability of the proteinaceous raw material. The  
 CC sequences shown in ABZ78180-ABZ78236 represent genes encoding the  
 CC A. niger proteases of the invention.  
 XX SQ Sequence 3080 BP; 755 A; 803 C; 735 G; 787 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1,11e-77 Length: 3080  
 Score: 887.50 Matches: 223  
 Percent Similarity: 48.15% Conservative: 77  
 Best Local Similarity: 35.79% Mismatches: 191  
 Query Match: 30.74% Indels: 133  
 DB: 24 Gaps: 17  
 US-09-712-338-2\_COPY\_19\_555 (1-537) x ABZ78226 (1-3080)  
 QY 13 GlnLeuProLysAsnProThrGlyValLysThrLeuThrAla-----AsnAsnVal 30  
 Db 563 CAATTTCTCCCGAGCGGAGGCATCACTGTGCTCAAGTCCAGTTGCATGAGAATGTG 622  
 QY 31 ThrIleArgTyrLysGluProGlyAlaGlu----- 40  
 Db 623 ACTATTCTTTCAAGAGGCTGTGTGAGA-GTATCTAGAAATAGCTTTTATCTCGATGC 681  
 QY 41 -----GlyValCysGluThrProGlyValLysThrProGlyValLysSerTyrSergly 54  
 Db 682 CGTGTGATGTTCAGCTCGAATTTCCGAACACTACGCCGGGTGTCGATCTTATCGGGC 741  
 QY 55 TyrValAspThrSerPro----- 60  
 Db 742 TATGTACACTTCCCGCCCGTTTCCTTCCGAGGAGACAGAGAGTATGATTAATCT 801

QY 61 -----Glu-SerHisTh 64  
 Db 802 ATCAACAGTAAGCAATCTCGAACAATGAGGATGAGCAATATACATGAGCCTCAACAG 861  
 QY 64 rPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLe 84  
 Db 862 CTTCCTTTGGTTTTTCGAAGCCGCCAAGATCCAGCAATCGCGCTCTGGCCANCTGGCT 921  
 QY 84 uAsnGlyClyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHi 104  
 Db 922 CAATGGCGGTCGGGTGGCTCGCTCGCTATGGGCTCTTGAAGAATAGTCTGTTTC 981  
 QY 104 sValAsnSer----ThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAs 123  
 Db 982 CATTGCCATCAGACTCCAGACCACAGCTCCATCTTGGAGTGGAAACAATGAGTCAA 1041  
 QY 123 nLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSer----- 138  
 Db 1042 TCTTCTATTCTTGACACGCCAACTCAAGTCGGCTTCTCATACGATGTCGCCAACAAATGG 1101  
 QY 139 -----AspThrValAspGlySerIleAsnProValThrGly-----Va 151  
 Db 1102 CACTTTGTCGACTCGGACGCGAAGAGATAGTTTCGGTGATTTCTCCATGA 1161  
 QY 151 lValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLe 171  
 Db 1162 TGTTCGCCAGTCCAACTTCACCCATCATGTGTGTACCTTTGCCAAGC-----CAGAAGCT 1215  
 QY 171 uIleAspThrAsn-----LeuAlaGluAlaAlaTrpGluIleLeuGlnG 188  
 Db 1216 TGCACAGACAGCTAAATGGGACTCGAATCGCGGCTCAGCTCTATGGCATTTTCGGCAAC 1275  
 QY 188 yPheLeuSerGlyLeuProSerLeu-----AspSerArgValGlnSerLysAspPh 205  
 Db 1276 CTGGTTTTTCAGTTCCACACATACAGCCAAAGCATGATCGTGTCTC----- 1321  
 QY 205 eSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTy 225  
 Db 1322 -AGTCTCTGGCTGAAATACGAGGCCATATGTCACAGGCATCTTCGGTTCCTTCCA 1380  
 QY 225 rGluGluAsnGluArgIleAlaAsnGlySerVal---AsnGlyValGln---LeuAsnPh 243  
 Db 1381 ACAGCAGAATGCAAAATCGCAGAGGGAGCTGCAGAAGACGGTGCACAGTATTTGCATCT 1440  
 QY 243 eAsnSerLeuGlyIleLeuAsnGlyIleLeuAspGluAlaIleGlnAlaProTyrTyrPr 263  
 Db 1441 CGACAGCTTGGCATTTGTAACGGCTTGATGGATATGGTATCCAAAGAGAGGCTTACAT 1500  
 QY 263 oGluPheAlaValAsnAsn----- 269  
 Db 1501 TACTTGGCCATACAACTAAGCTAAGCTCGCCCTTCTTCATTCACCTCGCTAATGCCVA 1560  
 QY 270 -----ThrTyrGlyIleLeuAlaValAsnGluThrValTyrAsnTyrMetLysP 286  
 Db 1561 ATTCAGTTCCAGACTACGCGCTCGAAATCTCGATAAACCCCTCTACGAAGAAGTATGAT 1620  
 QY 286 heAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnA 306  
 Db 1621 ATAACCTGGACCATCCAGGAGC----- 1643  
 QY 306 rgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp--- 324  
 Db 1644 -----TTTCGATCAGGCCCTCGCTCGCAAGCGGCTTGAAGAAGACGGATTCG 1695  
 QY 325 -----AsnVal----- 326  
 Db 1696 GCTTGCTCACTCAGGGAAGATATCTCGAAATTTGCGAGGCGCTTGCACATGAGATGG 1755  
 QY 327 --GluGlyPro-----TyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisP 344  
 Db 1756 GAGATGGCCCATCACCTACTACACACCTTCAATCGCGGGTGTACGACATCGGCCATC 1815

QY 344 rOIYrAspAspProThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValM 364  
 Db 1816 CTAGAAGACGCCCATTCCTCTCCAAAGCATCTCGGATATTTGACGAGGAGTCCGTCC 1875  
 QY 364 eAspAlaIleGlyValAsnIleAsnTyrThrGluSerAsnAsnAspValTyrTyrAlaP 384  
 Db 1876 TTGCGGCTCTGGGTACAGTCAATTCACAICCTCTTCGATCTCGGTCGGTGCACAGT 1935  
 QY 384 heGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIleLeuA 404  
 Db 1936 TCATAAAAACCTTTGATATCGTCCACGCGGCTTCCTGTGATGCAATTGCTACCTCC 1995  
 QY 404 laLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheG 424  
 Db 1996 ACATGGTGTAAAAAGTACACATGATGATACGGAGATCGTGATTACGCTGCAATGGT 2055  
 QY 424 lyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaG 444  
 Db 2056 GGGGGAANAACCCAGCCTTCAGTTCGATATTCGCTATTCACGGAATTCGCGACACGG 2115  
 QY 444 lyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsn 464  
 Db 2116 GATATCTCCATCTTACGCGGATCAGCGGATCAGCGCATCAGCGCCAGCTGGGCACT 2175  
 QY 464 heSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAla 484  
 Db 2176 ACAGCTTCACTCGCGCTTCCAAAGCGGCAAGGTCCTCTCTACAGCTGTGGGG 2235  
 QY 484 erLeuGlnLeuPheAsnArgThrIlePheGlyTyrAspIleAlaGluGlyGlnLysLys 504  
 Db 2236 CGIATGAGATCTTCATCGGCGGCACATTCACAAGATATCCCTACTGGCTCTTGGCTG 2295  
 QY 504 leTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSerValP 524  
 Db 2296 TTGATGACGAATTCAGTCTGGTTCACCTAAGGATACGTGSCATATCAAGATATCCCTC 2355  
 QY 524 rOleu 525  
 Db 2356 CTATT 2360  
 RESULT 10  
 ABZ51997  
 ID ABZ51997 standard; cDNA; 1007 bp.  
 XX ABZ51997;  
 XX 28-MAR-2003 (first entry)  
 XX Aspergillus oryzae polynucleotide SEQ ID NO 1110.  
 DE Aspergillus oryzae; fermentation; fungus; industrial; EST;  
 KW expressed sequence tag; gene; ss.  
 OS Aspergillus oryzae.  
 PN W0200279476-A1.  
 XX 10-OCT-2002.  
 XX 22-MAR-2002; 2002WO-IB00890.  
 XX 30-MAR-2001; 2001JP-0098371.  
 XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
 PA (NARE-) NAT RES INST BREWING.  
 PA (NOR) NAT FOOD RES INST MIN AGRIC.  
 XX Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;  
 PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;  
 DR WPI; 2003-046817/04.  
 XX  
 PT Detection of expression of specific Aspergillus genes for monitoring

the fermentation and growth conditions of the fungus, using DNA probes

Claim 1; SEQ ID NO 1110; 48pp + Sequence Listing; Japanese.

The invention relates to a polynucleotide having any of 6006 specific sequences (AB250888-AB256893), which are expressed by a fungus under specific culture conditions including one or more of eutrophic, oligotrophic, solid, early germination, alkaline, high temperature, low temperature or maltose culture or polynucleotides stringently hybridizing to these sequences. The polynucleotides are useful for monitoring the progress of fermentation and the growth conditions of a fungus, especially of *Aspergillus oryzae* which is widely used in industrial fermentation. Also monitoring for fungal contamination.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1007 BP; 253 A; 270 C; 257 G; 224 T; 3 other;

Alignment Scores:  
Pred. No.: 3,42e-60 Length: 1007  
Score: 704.50 Matches: 156  
Percent Similarity: 61.36% Conservative: 52  
Best Local Similarity: 46.02% Mismatches: 97  
Query Match: 24.40% Indels: 36  
DB: 25 Gaps: 9

US-09-712-338-2\_COPY\_19\_555 (1-537) x AB251997 (1-1007)

197 AspSerArgValGlnSerLysAspPheSerLeuThrTrpThrGlyGlyHisTyr 216  
18 GATCAGGAGTAC-----AGCATCTGGACGGAGC--TACGGAGGACGGTAC 60  
217 GlyProAlaPhePheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerVal 236  
61 GGGCATCTTTCATCTGCTTCTCCAGGAGCAGAAUGAGAGATGTGCCACGS-TCCATT 119  
237 -----AsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIle 253  
120 GACATCGACGATGCTCACTATATACCTGGACACCTCGGAATATCAATGGTGGTGC 179  
254 AspGluAlaIleGlnAlaProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIle 273  
180 GATCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 239  
274 LysAlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGly 293  
240 GAGGCTATCAACAACTGTATATGATATGATGCGGATGGAGGCTTGGAGCAAGCCGGAGG 299  
294 CysGlnAspLeuIleSerThrCysLysGln----- 303  
300 TGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 359  
304 ---ThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCys 322  
360 GGCAACAATGAGACCGCT-----AACAAGCTCTGTGCGAAAGCAACAACACTACG 410  
323 ArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg 342  
411 AGCAACCAAGTAGAAGGCCCATATTCATCTGGACGAGGCTATTACGACATCTCG 470  
343 HisProTyrAspAspProThrProSerTyrAsnLysPheLeuAlaLysAspSer 362  
471 CATTTGACCCCGATCTTCCCTCCCTGCTACTTGTGTTCTTAACACGACACGTCG 530  
363 ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspValTyrTyr 382  
531 GTCCAAAGGAGCCCTCGGGTTCGGTTAAATTCACCGAATCTGTTGATAGCGTCTACAAT 590  
393 AlaPheGlnGlnThrGlyAspPheValTyrPro-----AsnPheIleGlu 397  
591 GGCTTTTCCGCAACAGGTGACTAC-----CCACGCTCCGATGTACGGGGTACCTGGAG 644

398 AspLeuGluLurleuLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAsp 417  
645 GATATCGCATACGTCGTGACTCTGGCATCAAAAGTTGCCCTAGTGTATGCGATCGGAT 704  
418 TyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAla 437  
705 TAGCATGTCCTCGGACGAGGAGAGAGTGAAGTGAAGTGGAGTCTCGATGCC 764  
438 AlaGlnPheArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyr----- 455  
765 GCCAAGTTCGCTCTCGGTTTACGCCCTCTGAGACCAAT---GCCATATGATAGGT 821  
456 GlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGlu 475  
822 GGCTGTGTGCGACAGTWCAGAACTCTCGTTCCTACTCGGCTTTTCGAAGCGGTCAATGAG 881  
476 ValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTyr 495  
882 GCGCGCATATCAGCTGAACGGCGTATGAGATCTCCACCGAGCATGTGTTCAACAGA 941  
496 AspIleAlaGluGlyGlnLysLysIleTrpPro-----SerTyrLysThrAsnGly 512  
942 GACATTCGCGGCGCAGGCTCAATCGCCAAAGAACACACCTACTCCACTCAGCGG 998  
RESULT 11  
ABQ76315  
ID ABQ76315 standard; cDNA: 2027 BP.  
XX AC ABQ76315;  
XX DT 21-NOV-2002 (first entry)  
XX DE S. cerevisiae BAX-associated cDNA fragment SEQ ID 55.  
XX KW Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;  
XX KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;  
XX KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;  
XX KW neurodegeneration; cell death; ss.  
XX OS Saccharomyces cerevisiae.  
XX PN WO200264766-A2.  
XX PD 22-AUG-2002.  
XX PF 21-DEC-2001; 2001WO-EP15398.  
XX PR 22-DEC-2000; 2000EP-0870318.  
XX PR 04-JAN-2001; 2001EP-0870002.  
XX PR 09-JAN-2001; 2001EP-0870003.  
XX PA (JANC ) JANSSEN PHARM NV.  
XX PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;  
XX WPI; 2002-667002/71.  
XX P-PSDB; ABG93049.  
XX PT New isolated nucleic acid representing a synthetic Bax gene, useful as  
PT medicament for treating, preventing and/or alleviating yeast or fungal  
PT infections or proliferative disorders, or for preventing apoptosis in  
XX certain diseases -  
PS Claim 36; Figure 1; 344bp; English.  
XX This invention describes a novel nucleic acid representing a synthetic  
CC Bax gene. The Bax gene of the invention is useful for identifying  
CC Bax-resistant yeast or fungi, identifying, or obtaining and identifying  
CC Candida spp. sequences that are differentially expressed in a pathway  
CC eventually leading to programmed cell death or identifying inhibitors or  
CC inhibitor sequences of Bax-induced cell death. The products of the  
CC invention have cytostatic, fungicide; immunosuppressive, virucide and

CC vasotropic activity and can be used in vaccines or for gene therapy. The  
 CC isolated nucleic acids, polypeptides, pharmaceutical compositions,  
 CC antisense molecules and antibodies are useful as medicaments or in  
 CC preparing a medicament for treating, preventing and/or alleviating  
 CC diseases associated with yeast or fungi or proliferative disorders, such  
 CC as cancer, or for preventing apoptosis in certain diseases. The compounds  
 CC or polypeptides, or the genetically modified organism are useful for  
 CC preparing a medicament for modifying the endogenous flora of humans and  
 CC other mammals. The vaccine is useful for immunising against yeast or  
 CC fungal infections. Apoptosis-related diseases include autoimmune disease,  
 CC ischaemia, diseases related with viral infections or neurodegenerations.  
 CC This sequence represents a polynucleotide associated with the Bax gene  
 CC described in the disclosure of the invention.

XX  
 SQ sequence 2027 BP; 583 A; 380 C; 427 G; 637 T; 0 other;

## Alignment Scores:

Pred. No.:	5.76e-53	Length:	2027
Score:	636.00	Matches:	167
Percent Similarity:	46.29%	Conservative:	70
Best Local Similarity:	32.62%	Mismatches:	197
Query Match:	22.03%	Indels:	78
DB:	24	Gaps:	14

US-09-712-338-2\_COPY\_19\_555 (1-537) x ABQ76315 (1-2027)

```

QY 1 LeuProGlySerThrProAlaSerValGlyArgGlnLeu-----ProLys 16
D 627 CTTCCAGAAATACACGAAACCTTAAATGGACCGTTTGAATCAGGATGATCCCGCTG 686
QY 17 AsnProThrGlyValLysThrLeuThrThraAlaAsnAsnValThrIleArgTyrLysGlu 36
D 687 TTTTACAACCTTTATTTCTCTGTTGGACACAGATTACAGTTTGACATTAGACAGTAGAT 746
QY 37 ProGlyAlaGluGlyValCysGluThrProGlyValLysSerTyrSerGlyTyrVal 56
D 747 CTTCTTAAACTAGGAATT-----GACACCCGTAATAAATGGTCGGGTTCATG 794
QY 57 AspThrSerProGluSerHisThrPhePheThrPheGluAlaArgHisAsnProGlu 76
D 795 GACTATAAGGATTCACAAACACTTTTCTGTTAAATGGTGACCTGTTTCCCTGTTTACTGGTGG 854
QY 77 ThrAlaProIleThrLeuThrLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeu 96
D 855 AAGCAGCCCAATATCTTCTGTTAAATGGTGACCTGTTTCTGTTTACTGGTGGTGG 914
QY 97 PheGluGluLeuGlyProCysHisValAsnSerThrPheAspSerTyrIleAsnProHis 116
D 915 CTATTTGAACACTAGGCCCTCATCAATGTCGCGCGATATGAACCAATCCCAATCCCTAT 974
QY 117 SerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer 136
D 975 TCTTGGAAATAATACCGTTCATGATCTCTTGAACAGCCACTCGGAGTCGGCTTTTCC 1034
QY 137 TyrSerAspThrValAspGlySerIleAsnProValThrGlyValGluAsnSerSer 156
D 1035 TATGGTGAT----- 1043
QY 157 PheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsn 176
D 1044 -----GAAAAAGTCTCCTCTACAAAA 1064
QY 177 LeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeu 196
D 1065 TTAGCAGGCAAGATCGGTACATTTTCCGGAATTTGTTTTCGAAGCTTTTCTCATTTA 1124
QY 197 AspSerArgValGlnSerLysAspPheSerLeuThrPheGluSerTyrGlyGlyHisTyr 216
D 1125 -----CGCTCCCAAGATTTCCCAATTCAGGCGCAATCCCTATCGAGACATAT 1172
QY 217 GlyProAlaPhePheAsnHisPheTyrGluGlnAsn---GluArgIleAlaAsnGlySer 235
D 1173 ATCCCTCAAAATTCAGATCGTTGTGCAAGAACCTTGAAGA----- 1217

```

```

QY 236 ValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleLeuAspGlu 255
D 1218 -----ACGTTCAATTTAAGTTTCAATTTAAGTTTCAATTTAAGTTTCAATTTAAGTTT 1265
QY 256 AlaIleGlnAlaProTyrTyrProGluPheAlaVal---AsnAsnThrTyrGlyIleLys 274
D 1266 TTGATTCAGCAGATTTATTAACCAATGGCTCGCGGAAAGGGGGGTATCACCTGTT 1325
QY 275 AlaValAsnGluIleValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCys 294
D 1326 CTCCTCATCAGAAGAATGTGAGAAAATGAGTAAAGCTGCAGGTCG----- 1370
QY 295 GlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu 314
D 1371 -----TGCTGTAGTTGAACAAGTTATGTTATGTTCTTAAATCAAGT 1412
QY 315 -----CysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAla 332
D 1413 TTACATGCAATAGTCCGACCTGCTTACTGTGACTCIGACCTTTTGGAAACCGTACATTAAC 1472
QY 333 PheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro----- 348
D 1473 ---ACAGGACTCAACGTTCTATGACATAGAGGCGCTCTGAGATATAGTACTACTGATGTT 1529
QY 349 -----ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAsp 365
D 1530 ATGTGTTATACAGGTCTCCGCTATGTCAGCAGTATATGAATTTTCTGGAAGTTCAAGAA 1589
QY 366 AlaIleGlyValAsnIle---AsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPhe 384
D 1590 ACGTAGGTCGCGAGTGATTAATATTTCTGGCTGTGATATGATGATGTTCCACCGGATTT 1649
QY 385 GlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluGluIleLeuAla 404
D 1650 ITGTTTACGGCGGATGGAAGTAAACCA---TTTCAACAATATATGCTGAATTTAAAT 1706
QY 405 LeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGly 424
D 1707 CACAACATTCGGTATTATTAATATATGCGGTGATAGGATTAATTTCTGTAATTTGCTGG 1766
QY 425 GlyGlnAlaValSerLeuAlaAlaAspTyrSerGlnAlaGlnPheArgSerAlaGly 444
D 1767 AACCATGCTTGGTCCCAATGAGTTGGAATGGATCAATAAACAGTATGATCAGAGAAGGATG 1826
QY 445 TyrThrPro-----LeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 462
D 1827 TTAAGACCATGGGTCAGTAAAGAAACAGSTCAAGAGTGGGACAAAGTCAAGAACTATGGC 1886
QY 463 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 482
D 1887 CCTTTCACCTTTTGTAGAAATACGATGCGGTGATATGTTGCTGCTATGATCAACCGGAG 1946
QY 483 AlaSerLeuGlnLeuPheAsnArgThrIlePheGly 494
D 1947 GCAAGTTGGAATGGTCAACAGTTGGATTTCCGGT 1982

```

RESULT 12

ABZ78243

ID ABZ78243 standard; cDNA; 1611 BP.

XX ABZ78243;

AC ABZ78243;

DT 24-APR-2003 (first entry)

XX A. niger carboxypeptidase Y cDNA #1.

DE A. niger carboxypeptidase Y cDNA #1.

XX Protease; fungal infection; aspergillosis; food; tanning; detergent;  
 KW protein solubility; viscosity; taste; texture; nutritional value;  
 XX gene; ss.

OS Aspergillus niger.

XX



QY		266	AlaValAsnAsnThrTyrGlyLeuLysAlaValAsnGluThrValTyrAsnTyrMetLys	285
Db		913	GCC TGGGT GACGGCGGT ACCAGCTCTTGACGAGAGACTCCTGCCAGCTCATGCAC	972
QY		286	PheAlaasnGlnMetProAsnGlyCysGlnAspLeuSerThrCysLysGlnThrAsn	305
Db		973	-----AACGCTCTTCCTCGC-----TGCAGCTCTATGATTGAGTCTGCIACAGTCCGAG	1023
QY		306	ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsn	325
Db		1024	AGCGCT-----TGGTTTGTCCTCCGCGCTCACTACTCTAAACAAGGCC	1068
QY		326	ValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleAeqHisProTyr	345
Db		1069	CTCCTTGCCCTTACCAGCG-----ACTGGCGAGAACGCTCTATGAIGTCCGGTGGTAAGTGC	1125
QY		346	AspAspPro-----ThrProSerTyrTyrAsnLysPheLeuAlaLys	360
Db		1126	GAGTAGAGCTCTAACTTTGCTACTCGCTATGGCTACGTCACGCACTACTCTGAACAAG	1185
QY		361	AspSerValMetAspAlaIleGlyValAsnIleAsn--TyrThrGlnSerAsnAsnAsp	379
Db		1186	CCCCAAGTCAICGAGGCTGTGGCGCTGAGGTC AACGGCTACGACTCGTCCAACITTGAC	1245
QY		380	ValTyrTyrAlaPheGlnGlnThrGlyAspPheValTrpPro-----AsnPheIleGlu	397
Db		1246	ATCAACCGCAACTTCCTCTCCACGGTGACTGGATGAGCCCTACACCGCTCGTTCGG	1305
QY		398	AspLeuGluIleLeuAlaLeuProValargValSerLeuIleTyr---GlyAspAla	416
Db		1306	GGACTCTGTGAG-----CAGATCCCTGTC-----TTGATCTATGCCGTGATGCT	1350
QY		417	AspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGln	436
Db		1351	GAITTCATTGCACACTGCGCGGCAACAGGCCCTGGACTGAAGCCCTGGAGTGGCCCGGA	1410
QY		437	AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn	451
Db		1411	CAGGCTCAATGCTCCGCTGAGCTGGAGGATCTGGTCATTTGACATGAGCACACG	1470
QY		452	GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu	471
Db		1471	GGCAAGAAGATGGCCAGGTTAAGTCCCATGGCAACTTCACCTTCATGGCTCTCTATGGI	1530
QY		472	AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr	491
Db		1531	GGTGGCCACATGTTCCGATGGACCGCCAGCTCGAGICTCGAGTCTTTCACCCGCTGG	1590
QY		492	IlePheGly---Trp	495
Db		1591	TGCGGAGGTGAATGG	1605

## RESULT 13

RESULTS	
RESOLLI 13	
AAT28284	
ID	AAT28284 standard; cDNA; 2002 BP.
XX	
XX	
AC	AAT28284;
XX	
XX	
DT	14-AUG-1996 (first entry)
XX	
DE	A. niger SFAG 2 carboxypeptidase Y cDNA.
XX	
XX	
KW	Carboxypeptidase Y; CPY; ascomycete; deuteromycete; host cell;
KW	protease deficiency; ds.
XX	
OS	Aspergillus niger strain SFAG 2.

	key	Location/Qualifiers
FH	CDS	140..1813
FT		/*tag= a
FT		140..1810
FT	mat_peptide	/*tag= b
FT		

XX	WO9609397-AL.	
PN		
XX	28-MAR-1996.	
PD		
XX		
PF	19-SEP-1995;	95WO-US11945.
XX		
XX	20-SEP-1994;	94US-0309341.
PR		
XX		
PA	(NOVO ) NOVO NORDISK BIOTECH INC.	
XX		
PI	Thompson SA, Yaver DS;	
XX		
DR	WPI; 1996-188458/19.	
DR	P-ESDB; AAR96738.	
XX		
PT	Nucleic acid construct encoding a filamentous ascomycete or	
PT	deuteromycete carboxy:peptidase Y - useful to produce host cells	
PT	modified to produce reduced amounts of carboxy:peptidase	
XX		
PS	Claim 5; Page 23-25; 46pp; English.	
XX		
CC	The CPY gene (AAT28284) of Aspergillus niger strain SFAG 2 codes for	
CC	carboxypeptidase Y pre-propeptide (AAR96738). It was obtd. by	
CC	screening a cDNA library in Lambda ZAPII with a probe obtd. by PCR	
CC	amplification of A. niger Bo-1 genomic DNA (see also AAT28283).	
CC	The gene is useful in the creation of CPY-deficient mutants of	
CC	Aspergillus. For example, a selectable marker may be cloned into the	
CC	middle of the CPY gene. The resulting CPY-deficient mutant is useful	
CC	as a host for prodn. of heterologous protein.	
XX		
SQ	Sequence 2002 BP: 416 A; 590 C; 506 G; 490 T; 0 other;	

Alignment Scores:  
Pred. No.:  
Score:  
Percent Similarity:  
Best Local Similarity:  
Query Match:  
DB:

[illegible]

872 -----GTCAGCAGACCGTGTCTGCTGGCAGGACGGTCTATGCGCTTG 913  
 186 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 205  
 914 CTTACCTCTCTTCAACAATCCCG-----GAGTATGCCAAGCAGGACTTC 961  
 206 SerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsnHisPheTyr 225  
 962 CATATCCCGGTGAUCCATGCTATGCTACATATCCCGCTTTGCTGGAGATTTC 1021  
 226 GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 245  
 1022 TCTCACAGAAGCGC-----AACATCAACCTCCAGTCC 1054  
 246 LeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPhe 265  
 1055 GTTCTTATGGCAACGGCTCTCACCGAGCGTCTCACTGAGTACGAGTACTACCGTCCCATG 1114  
 266 AlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLys 285  
 1115 GCCTGTGTGACGGT---GGTATCCACGCTCTTGGACGAG---GGCTCTGCCAGGCC 1168  
 286 PheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsn 305  
 1169 ATGGACAACGCCCTCTCCGCG---TGCCAGTCTATGATGATGCTGCTATAGTTCGCGAG 1225  
 306 ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsn 325  
 1226 AGCGCT-----TGGGTTGTGTCCCGCGCTCCATCTACTGTAACAGCGCC 1270  
 326 ValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyr 345  
 1271 CTCCTTGCCCTTACCAGCG---ACCGACAGACACCTCTACGATGTTCTGGTAAGTGC 1327  
 346 AspAspPro-----ThrProSerTyrTyrAsnLysPheLeuAlaLys 360  
 1328 GAGGATAGCTCCAACTCTGCTACTCGCCATGGGCTACGTCACGCGACTACTCAACAAC 1387  
 361 AspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsp 379  
 1388 ACCGAGGTCATGAGGCTGTGGCGCTGAGGTCAACGCTACGACTCGTGCACATTTGAC 1447  
 380 ValTyrTyrAlaPheGlnGluThrGlyAspPheValTrpPro-----AsnPheIleGlu 397  
 1448 ATCAACCCCACTTCTCTCCACGGTGACTGGATGAGCCCTACCCCGTCTCGTCCG 1507  
 398 AspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAla 416  
 1508 GGACTCTCTGGAG-----CAGATCCCTGTC-----CTGACTACGCTGGTGAGGCC 1552  
 417 AspTyrIleCysAsnTrpPheGlyGlyClnAlaValSerLeuAlaAlaAsnTyrSerGln 436  
 1553 GATTTCATCTCACTGCTGGCAACAGCGCTGAGCTGAAGCCCTTGACTGGCGCGGA 1612  
 437 AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn 451  
 1613 CAGGCTGATATGCTCCGCTAAGCTGGAGACCTGGTGGTGGAGAACGACACAAG 1672  
 452 GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu 471  
 1673 GCAAGAAGATCGGCGCAGGTCAAGTCCCATGCCAATTCACCTTCATGCGTCTATGGC 1732  
 472 AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr 491  
 1733 GGTGGCCACATGGTCCCGATGACCAACCGAGTCTGAATCTTCAACCGCGTG 1792  
 492 IlePheGly---Trp 495  
 1793 TTGGGAGGTAATGG 1807  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 528  
 529  
 530  
 531  
 532  
 533  
 534  
 535  
 536  
 537  
 538  
 539  
 540  
 541  
 542  
 543  
 544  
 545  
 546  
 547  
 548  
 549  
 550  
 551  
 552  
 553  
 554  
 555  
 556  
 557  
 558  
 559  
 560  
 561  
 562  
 563  
 564  
 565  
 566  
 567  
 568  
 569  
 570  
 571  
 572  
 573  
 574  
 575  
 576  
 577  
 578  
 579  
 580  
 581  
 582  
 583  
 584  
 585  
 586  
 587  
 588  
 589  
 590  
 591  
 592  
 593  
 594  
 595  
 596  
 597  
 598  
 599  
 600  
 601  
 602  
 603  
 604  
 605  
 606  
 607  
 608  
 609  
 610  
 611  
 612  
 613  
 614  
 615  
 616  
 617  
 618  
 619  
 620  
 621  
 622  
 623  
 624  
 625  
 626  
 627  
 628  
 629  
 630  
 631  
 632  
 633  
 634  
 635  
 636  
 637  
 638  
 639  
 640  
 641  
 642  
 643  
 644  
 645  
 646  
 647  
 648  
 649  
 650  
 651  
 652  
 653  
 654  
 655  
 656  
 657  
 658  
 659  
 660  
 661  
 662  
 663  
 664  
 665  
 666  
 667  
 668  
 669  
 670  
 671  
 672  
 673  
 674  
 675  
 676  
 677  
 678  
 679  
 680  
 681  
 682  
 683  
 684  
 685  
 686  
 687  
 688  
 689  
 690  
 691  
 692  
 693  
 694  
 695  
 696  
 697  
 698  
 699  
 700  
 701  
 702  
 703  
 704  
 705  
 706  
 707  
 708  
 709  
 710  
 711  
 712  
 713  
 714  
 715  
 716  
 717  
 718  
 719  
 720  
 721  
 722  
 723  
 724  
 725  
 726  
 727  
 728  
 729  
 730  
 731  
 732  
 733  
 734  
 735  
 736  
 737  
 738  
 739  
 740  
 741  
 742  
 743  
 744  
 745  
 746  
 747  
 748  
 749  
 750  
 751  
 752  
 753  
 754  
 755  
 756  
 757  
 758  
 759  
 760  
 761  
 762  
 763  
 764  
 765  
 766  
 767  
 768  
 769  
 770  
 771  
 772  
 773  
 774  
 775  
 776  
 777  
 778  
 779  
 780  
 781  
 782  
 783  
 784  
 785  
 786  
 787  
 788  
 789  
 790  
 791  
 792  
 793  
 794  
 795  
 796  
 797  
 798  
 799  
 800  
 801  
 802  
 803  
 804  
 805  
 806  
 807  
 808  
 809  
 810  
 811  
 812  
 813  
 814  
 815  
 816  
 817  
 818  
 819  
 820  
 821  
 822  
 823  
 824  
 825  
 826  
 827  
 828  
 829  
 830  
 831  
 832



```

Db 568 TACTCTAAGCACACACCTTGGTTCAAGATCACGCTTTTATAT-GCTCTGATATCTAACCC 626
QY 68 -----PheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeu 84
Db 627 AACTTAGGTCTTCGAGTCTCGCAATGACCCGAGATGATCCCGTTGTTCTTGGCTG 686
QY 85 AsnGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGlyProCysHis 104
Db 687 AACGGTGGCCCTGGGTCTCTCCCTCACCGGTCTTCATGAGAGCTTGGCCCTAGCAGC 746
QY 105 ValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu 124
Db 747 ATCAACAAGAAGATCCACCGCGTCTACAAIAGTACGCTTGAACCTCCAAACGCGCTG 806
QY 125 LeuPheLeuSerGlnProLeuGlyValGlyPheSerIleAsnProHisSerTrpAsnGluValSerAsnLeu 124
Db 807 ATCTTCCTTGACCAACGCTGCAATGCTGCTTACTCTCTACAGTAAGTCTGCT----- 857
QY 145 IleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyr 164
Db 857 ----- 857
QY 165 ProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaTrpGlu 184
Db 858 -----GTACGCGACACGCTGCTGCTGGCAAGGACGCTCTATGCC 896
QY 185 IleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAsp 204
Db 897 TTGCTTACCCTCTCTCAACAATTCCTC-----GAGTATGCTAAGCAGGAC 944
QY 205 PheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPhe 224
Db 945 TTCACATTCGGGTGAATCTTATGCTGCTCACTATATCCCGCTTCGCTCGAGATC 1004
QY 225 TyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsn 244
Db 1005 CTGCTCACAGAAGACGC-----AACATCAACCTGCAG 1037
QY 245 SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGlu 264
Db 1038 TCCGTTCTCATGGCACGGCTCTACCGACGGATACACCCAGTACTACGAGTACTACCTGCC 1097
QY 265 PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMet 284
Db 1098 ATGGCTGCGGTGACGCGGTATCCAGCTCTCTTGGACGAGAGCTCTGCCAGTCCAIG 1157
QY 285 LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr 304
Db 1158 GAC-----AACGCTCTCTCTCGC---TGCCAGTCTATGATGAGTCTGTGTACAGTTCC 1208
QY 305 AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp 324
Db 1209 GAGAGCGCT-----TGCGTTGTCTCCGCGCTCCATCTACTGTAAACAAC 1253
QY 325 AsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisPro 344
Db 1254 GCCTCTCTTGCCTTACCAGCGC---ACTGGCAGACGCTGCTATGATCGCGTGGTAAG 1310
QY 345 TyrAspAspPro-----ThrProProSerIleTyrTyrAsnLysPheLeuAla 359
Db 1311 TCGGAGGATAGCTCTAACCTTTGCTACTCGGCTACGCTACGCTACGCTACCTGAAC 1370
QY 360 LysAspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsn 378
Db 1371 AAGCCCAAGTCAATGAGGCTGTGGCGCTGAGGTCAACGGCTACGACTCGTGCAACTTT 1430
QY 379 AspValTyrTyrAlaPheGlnGlnThrGlyAspPheValTrpPro-----AsnPheIle 396
Db 1431 GACATCAACGCAACTCTCTCTTCCAGGCTGATGATGAGGCTTACCAACCGCTCGGTI 1490
QY 397 GluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAsp 415

```

---

```

Db 1491 CCGGGACTCTCTGGAG-----CAGATCCCGTGC-----TTGATCTATGCGCGTGAT 1535
QY 416 AlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSer 435
Db 1536 GCTGATTTTCATTTGCAACTGGCTGGSCAACAGGCTGGACCTGAGCTGAGAGTGGCC 1595
QY 436 GlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal----- 450
Db 1596 GGACAGGCTGAATATGCCCTCCGCTGAGCTGGAGGATCTGCTCATTTGCGACAATGAGCAC 1655
QY 451 AsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyr 470
Db 1656 ACGGCAAGCAAGATTTGGCCAGGTAAAGTCCCATGGCAACTTCACTTCAATGCTCTCAT 1715
QY 471 GluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 490
Db 1716 GGTGTGGCCACATGCTCCGATGGACCAAGCCGAGTCGAGTCTGAGTTCTTCAACCGC 1775
QY 491 ThrIlePheGly---Trp 495
Db 1776 TGGTTGGGAGGTGAATGG 1793

```

RESULT 15  
ABZ78186

ID ABZ78186 standard; DNA; 2660 BP.

AC ABZ78186;

DT XX

XX 24-APR-2003 (first entry)

DE A. niger carboxypeptidase Y gene #1.

KW Protease; fungal infection; aspergillosis; food; tanning; detergent;

KW protein solubility; viscosity; taste; texture; nutritional value;

gene; ds.

OS Aspergillus niger.

XX WO2000268623-A2.

XX 06-SEP-2002.

XX 22-FEB-2002; 2002WO-EP01984.

XX 23-FEB-2001; 2001EP-0200657.

PR 23-FEB-2001; 2001EP-0200658.

PR 23-FEB-2001; 2001EP-0200660.

PR 26-FEB-2001; 2001EP-0200706.

PR 26-FEB-2001; 2001EP-0200707.

PR 26-FEB-2001; 2001EP-0200708.

PR 26-FEB-2001; 2001EP-0200719.

PR 28-MAR-2001; 2001EP-0000075.

PR 28-MAR-2001; 2001EP-0000078.

PR 28-MAR-2001; 2001EP-0000080.

PR 28-MAR-2001; 2001EP-0000087.

PR 28-MAR-2001; 2001EP-0000088.

PR 21-MAY-2001; 2001EP-0000156.

PR 21-MAY-2001; 2001EP-0000159.

PR 21-MAY-2001; 2001EP-0000160.

PR 21-MAY-2001; 2001EP-0000162.

PR 21-MAY-2001; 2001EP-0000165.

PR 21-MAY-2001; 2001EP-0000166.

PR 21-MAY-2001; 2001EP-0000168.

PR 21-JUN-2001; 2001EP-0000240.

PR 21-JUN-2001; 2001EP-0000242.

PR 21-JUN-2001; 2001EP-0000244.

PR 21-JUN-2001; 2001EP-0000246.

PR 12-JUL-2001; 2001EP-0000280.

PR 12-JUL-2001; 2001EP-0000285.

PR 30-JUL-2001; 2001EP-0000323.

PR 30-JUL-2001; 2001EP-0000327.

PR 02-AUG-2001; 2001EP-0000341.

PR 02-AUG-2001; 2001EP-0000342.





*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 19:02:15 ; Search time 554 Seconds  
(without alignments)  
7835.192 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662  
Perfect score: 1608  
Sequence: 1 cttccaggaagtacaccggc.....gcattgccagtgtgtatg 1508

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 100

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :				
N_Geneseq_19Jun03.*				
1:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*			
2:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*			
3:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*			
4:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*			
5:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*			
6:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*			
7:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*			
8:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*			
9:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*			
10:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*			
11:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*			
12:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*			
13:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*			
14:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*			
15:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*			
16:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*			
17:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*			
18:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*			
19:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*			
20:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*			
21:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*			
22:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*			
23:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*			
24:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*			
25:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*			

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	1608	100.0	1662 19	AAV28620 A. oryzae ATCC2038

ALIGNMENTS

RESULT 1  
AAV28620

ID	AAV28620 standard; DNA; 1662 BP.
XX	
AC	AAV28620;
XX	
DT	27-AUG-1998 (first entry)
XX	
DE	A. oryzae ATCC20386 carboxypeptidase I DNA.
XX	
KW	Carboxypeptidase I; flavour improving agent; hydrolysate; proteinaceous;
KW	food industry; ss.
XX	
OS	Aspergillus oryzae.
XX	
PH	Key Location/Qualifiers
FT	1..1662
FT	/*tag= a
FT	/product= carboxypeptidase I
FT	/note= "partial coding sequence"
XX	
FN	WO9814599-A1.
XX	
PD	09-APR-1998.
XX	
PF	03-OCT-1997; 97WO-US17977.
XX	
PR	27-NOV-1996; 96US-0757534.
PR	04-OCT-1996; 96US-0726880.
XX	
PA	(NOVO ) NOVO NORDISK BIOTECH INC.
PA	(NOVO ) NOVO NORDISK AS.
XX	
PI	Berka R, Blinkovsky A, Brown K, Dambmann C, Golightly E;
PI	Klotz A, Mathisen TE, Rey M;
XX	
DR	WPI; 1998-240098/21.
DR	P-PSDB; AAW56099.
XX	
PT	Carboxypeptidase from Aspergillus oryzae - hydrolyses proteinaceous
PT	substrates, useful for improving flavour of foods
XX	
PS	Claim 2; Fig 3; 82pp; English.
XX	
CC	This DNA sequence encodes carboxypeptidase I from Aspergillus oryzae.
CC	This polypeptide has an optimal activity in the range of pH 3.0-7.5 at
CC	25 deg. C, optimal activity in the range of 55-60 deg. C at pH 4, and a
CC	residual activity of at least 65.5% after 30 minutes at pH 4.0 and
CC	60 deg. C. It also has the capacity to hydrolyse X from N-CBZ-Ala-X where
CC	N-CBZ is N-carbobenzyloxy and X is any amino acid. The carboxypeptidases
CC	can be used for obtaining hydrolysates (which can be enriched in free
CC	glutamic acid or peptide bound glutamic acid residues) from proteinaceous
CC	substrates. The carboxypeptidases can be used in flavour-improving
CC	compositions in the food industry. The products can also be used for the
CC	production of polypeptides free of carboxypeptidase activity.
XX	
SQ	Sequence 1662 BP; 396 A; 468 C; 398 G; 400 T; 0 other;

Query Match	100.0%;	Score 1608;	DB 19;	Length 1662;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1608;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	CTTCAGAGACTACACCGGGTCGTCGGTAGAGACAGCTACCAAGACCCACCGGG	60
Db	55	CTTCAGAGAGTACACCGGGTCGTCGGTAGAGAGAGTACCAAGACCCACCGGG	114
Qy	61	GTCAGAGACTCTTACACCGGCAACAATGTCACCATCGGTAACAAGAACCCGGGCAGAG	120
Db	115	GTCAGAGACTCTTACACCGGCAACAATGTCACCATCGGTAACAAGAACCCGGGCAGAG	174
Qy	121	GGCGTCTGCGAGACTACCCGGGGTGTCAAAATCTACTCTGATATGCGACACCTCTCC	180
Db	175	GGCGTCTGCGAGACTACCCGGGGTGTCAAAATCTACTCTGATATGCGACACCTCTCC	234
Qy	181	GAGTCCCATACCTTCTTGTTCTTTCGAGCCACAGATACCCAGAACTGACCTATC	240

Db 235 GAGTCCCATACCTCTTCTGTTCTTGAAGCCAGACATACCCAGAAACTGCACCTATC 294  
QY 241 ACATTGTGGTTGAATGTGGCCCTGGAAGCGATTCCTTTGATCGGTCTCTTCGAGACGTTG 300  
Db 295 ACATTGTGGTTGAATGTGGCCCTGGAAGCGATTCCTTTGATCGGTCTCTTCGAGAGATTG 354  
QY 301 GGCCCTTGCCATGTCATTCGACTTTTGAATGACATACATCAACCCCTCACTCGTGAACGAG 360  
Db 355 GGCCCTTGCCATGTCATTCGACTTTTGAATGACATACATCAACCCCTCACTCGTGAACGAG 414  
QY 361 GTCTCCAAATTTACTATTCCTGTCGCCAGCCATTTGGAGTCGGCTTTTTCATATAGTATACG 420  
Db 415 GTCTCCAAATTTACTATTCCTGTCGCCAGCCATTTGGAGTCGGCTTTTTCATATAGTATACG 474  
QY 421 GTTATGGGTCCCAATTAACCTGTAACCTGGGTCTGCAAAATTCGAGCTTTGCGAGGATT 480  
Db 475 GTTATGGGTCCCAATTAACCTGTAACCTGGGTCTGCAAAATTCGAGCTTTGCGAGGATT 534  
QY 481 CAGGGCCGGTACCCAAACCATTTGATGCCACTCTGATCGATACCAATCTTTGCCGACAGAG 540  
Db 535 CAGGGCCGGTACCCAAACCATTTGATGCCACTCTGATCGATACCAATCTTTGCCGACAGAG 594  
QY 541 GCCGCTTGGGAGATCCCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCAGGGTG 600  
Db 595 GCCGCTTGGGAGATCCCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCAGGGTG 654  
QY 601 CAGTCTAAGGACTTCAGTCTATGACGAGAGACTATGGAGGCACTATGCTCTGCAATTC 660  
Db 655 CAGTCTAAGGACTTCAGTCTATGACGAGAGACTATGGAGGCACTATGCTCTGCAATTC 714  
QY 661 TTCAATCATTTTACGAGCAGAAATGAGAGAAATGCCAACGGTAGTGTAAATGGTGTTCAG 720  
Db 715 TTCAATCATTTTACGAGCAGAAATGAGAGAAATGCCAACGGTAGTGTAAATGGTGTTCAG 774  
QY 721 CTTAATTTCACTCTCTGGGAATTTAAGCGGATCATCGAGGGGATCCAGGCCCT 780  
Db 775 CTTAATTTCACTCTCTGGGAATTTAAGCGGATCATCGAGGGGATCCAGGCCCT 834  
QY 781 TACTACCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACCGTC 840  
Db 835 TACTACCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACCGTC 894  
QY 841 TACAACCTACATGAAGTTTGCCAAACCAATGCCAAATGGTTGCCAGGATTTGATTTCCACC 900  
Db 895 TACAACCTACATGAAGTTTGCCAAACCAATGCCAAATGGTTGCCAGGATTTGATTTCCACC 954  
QY 901 TGCAACAGACAACCCAGCCGATTTAGCTGACTACGGCTCTCGCGCGAAGCCCAAC 960  
Db 955 TGCAACAGACAACCCAGCCGATTTAGCTGACTACGGCTCTCGCGCGAAGCCCAAC 1014  
QY 961 ATGTGCAGGGACAATTTGAGGGGCCATACCTACGCGCTTTGCTGGTGGTGGTGTATGAT 1020  
Db 1015 ATGTGCAGGGACAATTTGAGGGGCCATACCTACGCGCTTTGCTGGTGGTGGTGTATGAT 1074  
QY 1021 ATTGGGCATCCATATGATGACCCGACTCCGCAAGTTATTACAACAAATTTCTGGCAAG 1080  
Db 1075 ATTGGGCATCCATATGATGACCCGACTCCGCAAGTTATTACAACAAATTTCTGGCAAG 1134  
QY 1081 GACTCTGTCATGGAGCTATCGGGTCAACATCAACTACACCCAGTCCCAATTAATGACGTC 1140  
Db 1135 GACTCTGTCATGGAGCTATCGGGTCAACATCAACTACACCCAGTCCCAATTAATGACGTC 1194  
QY 1141 TACTACGCTTCCAGCAAAAGGCGACTTTGCTGGGCCCAACTTCATCGAAGAGCTCGAG 1200  
Db 1195 TACTACGCTTCCAGCAAAAGGCGACTTTGCTGGGCCCAACTTCATCGAAGAGCTCGAG 1254  
QY 1201 GAGATCCTTGCCTCCCGGTGCGTGTCTCCCTCATCTATGCGGACGCCGATTACATCTGC 1260  
Db 1255 GAGATCCTTGCCTCCCGGTGCGTGTCTCCCTCATCTATGCGGACGCCGATTACATCTGC 1314  
QY 1261 AACTGGTTTCGGGCTCAGGCGGTTTCCTGCTCGGAACCTACCTCCAGCGCGCCAGTTC 1320

Db 1315 AACTGGTTTCGGGCTCAGGCGGTTTCCTGCTCGTGAAGTACTCCCAAGCGCGCCAGTTC 1374  
QY 1321 CGAAGCGCAGGTTACACGCCCTCAAGAGTCAACGCGCTCGAGTATGGGCAAACTCGCGAG 1380  
Db 1375 CGAAGCGCAGGTTACACGCCCTCAAGAGTCAACGCGCTCGAGTATGGGCAAACTCGCGAG 1434  
QY 1381 TATGTAATTTCTCCTTCACTCGGCTCTATGAGSCAGGCCATGAAGTCCCATACTACACAG 1440  
Db 1435 TATGTAATTTCTCCTTCACTCGGCTCTATGAGSCAGGCCATGAAGTCCCATACTACACAG 1494  
QY 1441 CCATCGGCTCCCTGCAATTTTAAACCGGACTATCTTCGGTGGGATATCGAGAGGC 1500  
Db 1495 CCCATCGGCTCCCTGCAATTTTAAACCGGACTATCTTCGGTGGGATATCGAGAGGC 1554  
QY 1501 CAGAAGAGATCTGSCCAGCTACAAGACGAATGAACGGCTACAGCTACGATACACAG 1560  
Db 1555 CAGAAGAGATCTGSCCAGCTACAAGACGAATGAACGGCTACAGCTACGATACACAG 1614  
QY 1561 TCGTCCGTGCGCTGCGCTACGGCTACCGGCTACCGGCTACGCTGTGTGATG 1608  
Db 1615 TCGTCCGTGCGCTGCGCTACGGCTACCGGCTACCGGCTACGCTGTGTGATG 1662

Search completed: September 16, 2003, 19:15:52  
Job time : 554 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 14:42:53 ; Search time 554 seconds  
(without alignments)  
7835.192 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662

Perfect score: 1608

Sequence: 1 cttcaggaagtaacacggc.....gcattgcaggttggtatg 1608

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*

1:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1608	100.0	1662	19	AAV28620
2	874.8	54.4	1656	24	ABZ78288
3	313.6	19.5	3150	24	ABZ78231
4	164.8	10.2	1007	25	ABZ51997
5	134.4	8.4	1872	24	ABZ78289
6	118.6	7.4	1665	24	ABZ78283
7	116.8	7.3	1581	24	ABZ78241
8	116.8	7.3	2940	24	ABZ78184
					A. oryzae ATCC2038
					A. niger serine ca
					A. niger serine ca
					Aspergillus oryzae
					A. niger serine ca
					A. niger serine ca
					A. niger serine ca
					A. niger serine ca

9	114.6	7.1	3080	24	ABZ78226
10	107.8	6.7	3221	24	ABZ78232
11	103	6.4	678	21	AAF12116
12	87.8	5.5	636	21	AAF13804
13	86.4	5.4	2503	15	AAQ55347
14	84	5.2	2002	17	AAT28284
15	82.4	5.1	1611	24	ABZ78243
16	78.2	4.9	2068	17	AAI28283
17	78.2	4.9	2660	24	ABZ78186
18	75.6	4.7	626	21	AAF12522
19	68.4	4.3	1368	24	ABZ78267
20	68.4	4.3	1653	24	ABQ76547
21	68.4	4.3	2441	24	ABZ78210
22	63.2	3.9	1446	24	ABZ78269
23	61.8	3.8	2027	24	ABQ76315
24	61.8	3.8	1742	23	ABL07915
25	58.6	3.6	3826	23	ABL07914
26	58.6	3.6	3826	23	ABL07914
27	57.4	3.6	4509	23	ABL08118
28	57.4	3.6	475	25	ABZ54889
29	56.4	3.5	1814	22	AAF68401
30	56.4	3.5	1814	24	ABK38312
31	56.4	3.5	1814	25	ACA10641
32	56.4	3.5	1814	25	ABX99592
33	56.4	3.5	1815	24	ABQ88162
34	56.4	3.5	1815	24	ABN95798
35	54.8	3.4	2864	21	AAF18312
36	49.2	3.1	1997	25	ACC46113
37	48.4	3.0	1839	24	ABZ78258
38	48.4	3.0	2730	24	ABZ78201
39	48	3.0	869	21	AAF14932
40	47.6	3.0	390	13	AAQ21833
41	47.6	3.0	390	14	AAQ36859
42	47.6	3.0	390	22	AAF76910
43	47.6	3.0	390	24	AAI72775
44	47.6	3.0	1428	22	AAF94477
45	47.6	3.0	1551	19	AAV64076

## ALIGNMENTS

RESULT 1  
AAV28620  
ID AAV28620 standard; DNA; 1662 BP.  
XX  
AC AAV28620;  
XX  
DT 27-AUG-1998 (first entry)  
XX  
DE A. oryzae ATCC20386 carboxypeptidase I DNA.  
XX  
KW Carboxypeptidase I; flavour improving agent; hydrolase; proteinaceous;  
KW food industry; ss.  
XX  
OS Aspergillus oryzae.

XX Key Location/Qualifiers  
XX CDS 1..1662  
XX /\*tag= a  
XX /product= carboxypeptidase I  
XX /note= "partial coding sequence"

WO9814599-A1

09-APR-1998

03-OCT-1997

27-NOV-1996

04-OCT-1996

96US-0757534

96US-0726880

(NOVO ) NOVO NORDISK BIOTECH INC.

10262, 100%

PA (NOVO ) NOVO-NORDISK AS.  
XX Berka R, Blinkovsky A, Brown K, Dambmann C, Golightly E;  
PI Klotz A, Mathisen TE, Rey M;  
XX WPI; 1998-240098/21.  
DR P-PSDB; AAW56099.  
XX Carboxypeptidase from Aspergillus oryzae - hydrolyses proteinaceous  
PT substrates, useful for improving flavour of foods  
XX Claim 2; Fig 3; 82pp; English.  
XX This DNA sequence encodes carboxypeptidase I from Aspergillus oryzae.  
CC This polypeptide has an optimal activity in the range of pH 3.0-7.5 at  
CC 25 deg. C, optimal activity in the range of 55-60 deg. C at pH 4, and a  
CC residual activity of at least 65.5% after 30 minutes at pH 4.0 and  
CC 60 deg. C. It also has the capacity to hydrolyse X from N-Cbz-Ala-X where  
CC N-Cbz is N-carboxybenzoyl and X is any amino acid. The carboxypeptidases  
CC can be used for obtaining hydrolysates (which can be enriched in free  
CC glutamic acid or peptide bound glutamic acid residues) from proteinaceous  
CC substrates. The carboxypeptidases can be used in flavour-improving  
CC compositions in the food industry. The products can also be used for the  
CC production of polypeptides free of carboxypeptidase activity.  
XX Sequence 1662 BP; 396 A; 468 C; 398 G; 400 T; 0 other;  
Query Match 100.0%; Score 1608; DB 19; Length 1662;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CTTCCAGGAGTACACCGGGTCCGTCGGTAGAGACAGTACCCAGAACCCACCGGG 60  
DB 55 CTTCCAGGAGTACACCGGGTCCGTCGGTAGAGACAGTACCCAGAACCCACCGGG 114  
QY 61 GTCAGGACTCTTAAACCGCAAAACATGTCACCATCCGGTACAGGAACCCGGGACAG 120  
DB 115 GTCAGGACTCTTAAACCGCAAAACATGTCACCATCCGGTACAGGAACCCGGGACAG 174  
QY 121 GGGCTGCGAGACTACCGGGTCAAAATCTTCTGATATGTCGACCTCTCC 180  
DB 175 GGGCTGCGAGACTACCGGGTCAAAATCTTCTGATATGTCGACCTCTCC 234  
QY 181 GAGTCCCATACCTTCTTCTGTTCTTGAAGCCAGACATAACCCAGAACTGCACCTATC 240  
DB 235 GAGTCCCATACCTTCTTCTGTTCTTGAAGCCAGACATAACCCAGAACTGCACCTATC 294  
QY 241 ACATTGGTGTGAATGGTGGCCCTGGAGGCCATTTTGTATCGGTCTCTTGAAGAGTTG 300  
DB 295 ACATTGGTGTGAATGGTGGCCCTGGAGGCCATTTTGTATCGGTCTCTTGAAGAGTTG 354  
QY 301 GGGCTTCCCATGTCAATTCGACTTTTGTATGACTACATCAACCCCTCACTCGTGGAAACGAG 360  
DB 355 GGGCTTCCCATGTCAATTCGACTTTTGTATGACTACATCAACCCCTCACTCGTGGAAACGAG 414  
QY 361 GTCTCCAAATTAATTCCTCTCCAGCCATTTGGAGTGGGCTTTTCATATAGTATAGCAG 420  
DB 415 GTCTCCAAATTAATTCCTCTCCAGCCATTTGGAGTGGGCTTTTCATATAGTATAGCAG 474  
QY 421 GTTATGGGTTCATTAACCTTGTAATCGGGTCTCGAAATTCGAGCTTTTCAGAGTT 480  
DB 475 GTTATGGGTTCATTAACCTTGTAATCGGGTCTCGAAATTCGAGCTTTTCAGAGTT 534  
QY 481 CAGGCGCGGTACCCCAACCATTTGATGCCACTCTGATCGATACTACCAATTTTCGCCAGAG 540  
DB 535 CAGGCGCGGTACCCCAACCATTTGATGCCACTCTGATCGATACTACCAATTTTCGCCAGAG 594  
QY 541 GCGCGTTGGGAGATCTCAAGGATTCCTTATGTGACTACCTAGCTTGGACTCTAGGGTG 600  
DB 595 GCGCGTTGGGAGATCTCAAGGATTCCTTATGTGACTACCTAGCTTGGACTCTAGGGTG 654  
QY 601 CAGTCTAAGGACTTCAGTCTATGGACGGAGGACTATGGAGGSCACTATGGTCTCTGATTC 660  
|||||

RESULT 2  
ABZ78288  
ID ABZ78288 standard; cdna; 1656 BP.  
XX  
AC ABZ78288;  
XX

DB 655 CAGTCTAAGGACTTCAGTCTATGACGAGAGCTATGAGGAGCCTACTATGGTCTCGCATTC 714  
QY 661 TTCAATCATTTTTTACGAGCAGAAATGCCAAGATTCGCAACGCTAGCTGTATTAATGGTTTCAG 720  
DB 715 TTCAATCATTTTTTACGAGCAGAAATGCCAAGATTCGCAACGCTAGCTGTATTAATGGTTTCAG 774  
QY 721 CTTAATTTCAACTCTCTGGGAAATTTAATACGGCATCATPCGACGAGGAGGATCCAGGCGCCT 780  
DB 775 CTTAATTTCAACTCTCTGGGAAATTTAATACGGCATCATPCGACGAGGAGGATCCAGGCGCCT 834  
QY 781 TACTACCTGAAATTCGCTGTGAACAATACCTTCAAGGTATCAAGGCTGTCAACGACGACCTC 840  
DB 835 TACTACCTGAAATTCGCTGTGAACAATACCTTCAAGGTATCAAGGCTGTCAACGACGACCTC 894  
QY 841 TACAACATACATGAAGTTTGCAACCAATGTTCCCAATGTTCCAGGATTTGATTTCCACC 900  
DB 895 TACAACATACATGAAGTTTGCAACCAATGTTCCCAATGTTCCAGGATTTGATTTCCACC 954  
QY 901 TGAACACAGACAAACCCGACCGCATAGCTAGCTACGCTCTGCGGCGAAGCCACCAAC 960  
DB 955 TGAACACAGACAAACCCGACCGCATAGCTAGCTACGCTCTGCGGCGAAGCCACCAAC 1014  
QY 961 ATGTGCGAGGACAATGTTGAGGGCCATACCTACGCTTGTCTGCTGCTGCTGTGTATGAT 1020  
DB 1015 ATGTGCGAGGACAATGTTGAGGGCCATACCTACGCTTGTCTGCTGCTGCTGTGTATGAT 1074  
QY 1021 ATTTGGCATCCATATGATGACCCGACTCCGCAAGTTATTACAAATAATTTCTGGCAAG 1080  
DB 1075 ATTTGGCATCCATATGATGACCCGACTCCGCAAGTTATTACAAATAATTTCTGGCAAG 1134  
QY 1081 GACTCTGTATGAGGCTATCGGCTCAACATCACTACCCAGTCCCAATTAATACCTC 1140  
DB 1135 GACTCTGTATGAGGCTATCGGCTCAACATCACTACCCAGTCCCAATTAATACCTC 1194  
QY 1141 TACTACGCTTTCACGAAACAGGGGACTTTGTCTGGCCCACTTCATCGAAGACCTCGAG 1200  
DB 1195 TACTACGCTTTCACGAAACAGGGGACTTTGTCTGGCCCACTTCATCGAAGACCTCGAG 1254  
QY 1201 GAGATCCCTGTCTCCCGTGGTGTCTCCCTCATCTATGGCGAGCGGATTTACATCTGC 1260  
DB 1255 GAGATCCCTGTCTCCCGTGGTGTCTCCCTCATCTATGGCGAGCGGATTTACATCTGC 1314  
QY 1261 AACTGGTTCGGGGTTCAGGCGGTTTCCCTCGCTGGGAATCTACTCCAGCCGCCAGTTTC 1320  
DB 1315 AACTGGTTCGGGGTTCAGGCGGTTTCCCTCGCTGGGAATCTACTCCAGCCGCCAGTTTC 1374  
QY 1321 CGAAGCGAGGTTACACGCCCTGAAAGTCAACGGCGTCTGAGTATGGGAAATCTCGCGAG 1380  
DB 1375 CGAAGCGAGGTTACACGCCCTGAAAGTCAACGGCGTCTGAGTATGGGAAATCTCGCGAG 1434  
QY 1381 TATGTAATTTCTCTCACTCGGCTATGAGGAGGCGCATGAGTCCCATACTACAG 1440  
DB 1435 TATGTAATTTCTCTCACTCGGCTATGAGGAGGCGCATGAGTCCCATACTACAG 1494  
QY 1441 CCCATCGCTCCCTGCAATTTTAAACGGGACTATCTCGGTTGGATATCGCAGAGGC 1500  
DB 1495 CCCATCGCTCCCTGCAATTTTAAACGGGACTATCTTCGGTTGGATATCGCAGAGGC 1554  
QY 1501 CAGAAGAAGATCTGGCCAGCTTACAAGAGCAATGGAACGGCTACAGCTACGATACAG 1560  
DB 1555 CAGAAGAAGATCTGGCCAGCTTACAAGAGCAATGGAACGGCTACAGCTACGATACAG 1614  
QY 1561 TCGTCCGTCCCGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACG 1608  
DB 1615 TCGTCCGTCCCGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACG 1662







Claim 1; Page 187-189; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A polypeptide or polynucleotide of the invention is useful for diagnosing a fungal infection such as aspergillosis, or as a query sequence to perform a search against public databases. A polypeptide of the invention is useful in a selected number of industrial or pharmaceutical processes, in laboratory or clinical processes, in food industry (baking, brewing, cheese manufacturing, meat tenderising), in tanning industry and in the manufacture of biological detergents. A polypeptide may also be useful for improving protein solubility, extraction yields, viscosity or taste, texture, nutritional value, minimising of antigenicity or anti-nutritional factors, colour or functionality as well as processing aspects like filterability of the proteinaceous raw material. The sequences shown in AB278180-AB278236 represent genes encoding the A. niger proteases of the invention.

Sequence 3150 BP: 801 A; 778 C; 738 G; 833 T; 0 other.

Query Match	19.5%	Score 313.6	DB 24	Length 3150
Best Local Similarity	72.5%	Pred. No. 4.2e-83		
Matches 406	Conservative 0	Mismatches 154	Indels 0	Gaps 0
QY	1037	ATGACCGGACTCGCCCAAGTATTACAAACAAATTCCTGGCAAGGACTCTGTCATNGAAGC	1096	
Db	2194	AGGACCGGACCCGCGCGCTCTACTTTGTTGACATACCTCAAGAAAGACTCACTCATGGAIG	2253	
QY	1097	CTATCGCGCTCAACATCAACATACACCCAGTCCCAATAATGAGCTCTACTACGCTTTCACGC	1156	
Db	2254	CTATCGCGCTGGACATTAACTACACCGAGTCCACCGCGCGAAGTATATTGCAATCCACGC	2313	
QY	1157	AAACAGCGGACTTGTCTGGCCCACTTCATPCGAAGACCTCGAGGAGATCCTTGTCTCTCC	1216	
Db	2314	AGACCGCGGACTTTGATGCGCGAAATTCATTGAGGACCTCGAAGAGATGCTCCAACTCC	2373	
QY	1217	CCGTGGGTGTCTCCCTCATCTATGGGACGCCGATACATCTCAACTGCTGTTTCGGCGGTC	1276	
Db	2374	CCGTACGCTGTCTGTGATCTACGGGATGCCGACTATATCTGTAACCTGTTTCGGCGGTC	2433	
QY	1277	AGGCGGTTTCCCTPGCTGCGAATCTCTCCAAAGCCGCCCCAGTTCGGAAGCGCAGGGTACA	1336	
Db	2434	AGGCCATCTCACTCGGAGTTAACTACCCCATGCAGCTCAGTTCGCTGACGCGGATACA	2493	
QY	1337	CGSCCTCGAAGTCAACGGCGTCGAGTATCGGGAACCTCCGAGTAGTATGTTATTCCT	1396	
Db	2494	CACCCATGACAGTAGATGGGGTCCGAATACGGTGAGACTCCGAGTAGTATGGAACCTTCGT	2553	
QY	1397	TCACTCGGCTCTATGAGGCAGGCCATGAAGTCCCACTACACGCCCCATCGCCTCCCTGC	1456	
Db	2554	TCACCCGCTATATACGGCTGGSCACAGGTTCCATACATCAACCGATCGACGCTGC	2613	
QY	1457	AAATGTTTACCGGACTATCTTCGGTGTGGATATPCGAGAGGCCAGAGAAGATCTGGC	1516	
Db	2614	AGCTGTTTCAACCGTACTTTATTTGGATGGGATATTGACGGGGTACAACTTCAGATTGGC	2673	
QY	1517	CCAGCTACAAGACGAATGGAACGGCTACAGCTACGCATACACAGTCTCGTCCGCGCTGC	1576	
Db	2674	CCGAATATAGCACCAACGGGACATCCAGCTACACACAGGAGTGGTTCGTGCCACTGT	2733	
QY	1577	CTACGGCTACCGCATGTCTC	1596	
Db	2734	CCACGGCTCGAGTACCACC	2753	

RESULT 4  
AB251997  
ID AB251997 standard; cDNA; 1007 BP.  
XX  
XX AB251997;  
XX  
DT 28-MAR-2003 (first entry)

DE	Aspergillus oryzae polynucleotide SEQ ID NO 1110.	
XX		
KW	Aspergillus oryzae; fermentation; fungus; industrial; EST;	
KW	expressed sequence tag; gene; ss.	
XX		
OS	Aspergillus oryzae.	
PN	WO200279476-A1.	
XX		
PD	10-OCT-2002.	
XX		
PF	22-MAR-2002; 2002WO-IB00890.	
XX		
PR	30-MAR-2001; 2001JP-0098371.	
XX		
PA	(NNAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.	
PA	(NARE-) NAT RES INST BREWING.	
PA	(NORQ) NAT FOOD RES INST MIN AGRIC.	
XX		
PI	Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;	
PI	Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;	
XX		
DR	WPT; 2003-045817/04.	
XX		
PT	Detection of expression of specific Aspergillus genes for monitoring	
PT	the fermentation and growth conditions of the fungus, using DNA probes	
PT		
PS	Claim 1; SEQ ID NO 1110; 48pp + Sequence Listing; Japanese.	
XX		
CC	The invention relates to a polynucleotide having any of 6006 specific	
CC	sequences (ABZ5088-ABZ56893), which are expressed by a fungus under	
CC	specific culture conditions including one or more of eutrophic,	
CC	oligotrophic, solid, early germination, alkaline, high temperature, low	
CC	temperature or maltose culture or polynucleotides stringently hybridising	
CC	to these sequences. The polynucleotides are useful for monitoring the	
CC	progress of fermentation and the growth conditions of a fungus,	
CC	especially of Aspergillus oryzae which is widely used in industrial	
CC	fermentation. Also monitoring for fungal contamination.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published.pct sequences.	

[illegible]



Query Match	8.4%;	Score 134.4;	DB 24;	Length 1872;
Best Local Similarity	48.1%;	Prod. No. 2e-29;		
Matches 644;	Conservative 0;	Mismatches 661;	Indels 33;	Gaps 8;
QY	186	CCATAACCTCTCTCTCGTCTTCGGAAGCCAGACAT-----AACCCAGAAACGACACCTAT	239	
DB	282	CAATACCTTTTCTCGTCTCTCTTCGGAAGCCAGACAT-----AACCCAGAAACGACACCTAT	341	
QY	240	CACATTTGGTGTGAATGTGGCCCTGGAAGCGATCTTTTGATCGTCTCTTCGGAAGCTT	299	
DB	342	CACCAATCGGATGACGCGCGCGCGGATCTCTCAATGATTTGGCTATTTCAGAGAA	401	
QY	300	GGGCCCTTGCCATGTCAATTCGACHTTTGAJCACTACATCAACCCCTCAC---TCGTGGAA	356	
DB	402	CGGCCCATGTCTGTGAATACGGACTCGAAATCCACGCCCTAIAATCCCTGCTGCGAA	461	
QY	357	CGAGTCTCCCAATTTACTATTCTCTGCTCCACGCCATTGGAGTCTGGGATCTCTCATATGTA	416	
DB	462	TGAGTACGTGATATGTTGATATGACGACCGGTGCAGACGGGATTTAGTTATGATGT	522	
QY	417	TACGGTTGATGGGTCCATTAAACCCCTGTAACTGGGCTCGTGGAAATTCAGCTTTGCAGG	476	
DB	522	GTTCAGGAAATGGGACGTTAGAT---TTGAATGAGACGTTTGTGGGGACGTTGCCG-576		
QY	477	AGTTACGGCGCGTACCAACCATTTGATGCCACTCTGATCGAATCTACCAATCTTTCGCCG	536	
DB	577	-AGTCAGGATGTGATGGGACGGTGAATGGGACGGTTAATGGGGAAGGCGCTTGGGT	635	
QY	537	AGAGGCGCGCTGGGAGATCTCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACICTAG	596	
DB	636	TGCGTTGCAGGTTTGGTTGGGTGAATCTCTGAATATGTTTCTCTGTGTGACGGGAATGG	695	
QY	597	GGTCGAGCTTAAGGACTTTCAGTCTATGACCGGAGAGCTATGGAGGGCACTATGTCCTGC	656	
DB	696	TGTTGGTGTATGACAGGCTGAGTATATGACGAGCTCATATGGGACGGTATGACACCGC	755	
QY	657	ATTCCTTCAATCAATTTTACGAGCAGATGAGAGAATTCGCAACGGTAGTGT---TAATGG	713	
DB	756	ATACACGGCGCTCTTTACGAGAGTGAATGAGAGGATTCAGAGTGGGAGGTAAAGCACCGG	815	
QY	714	TGTTTCAGCTTAATTCAACTCTCTGGGAATTTATTAACGGCATCATGACGAGGCGATCCA	773	
DB	816	GAGAAGATCCATTTGATGATGCGTGGCAATTTCAATGGGTGTGTGATTTACTCTGTGCA	875	
QY	774	GGCCCTTACTACCTGAATTCGCTGTGAACAATACCTAGGGTATCAAGGCTGTCACACA	833	
DB	876	GGTCCCTTCGTTCCCTGACGAGCGGTATAACAATACGATATGGGATCGAGGAATCAATCG	935	
QY	834	GACGCTCTACAACCTACATGATAGTTTGCCACCAAAATGCCAAATGTTGGTCAGGATTTGAT	893	
DB	936	CAGCTCTACGACCGGGCTATGATAGTTTGAGCAAGCCTGGGGGTGTCAGGAGATATGAT	995	
QY	894	TTCACCTGCAAAACAGACAAACCGCACCGCATTTAGCTAGCTACGCCCTCTCGCCCGAGC	953	
DB	996	CATCGAGTGTGCGATGCTTGGCCAGCTCGG---AGATCCCTCATCATCTGCGAGGAGGC	1052	
QY	954	CACCAACATGTGAGGACAAATGTTGAGGGCCCATACTACGCCCTTTCTGCTGCTGCTGT	1013	
DB	1053	GTCCGACTACTGTTCCGCGGAGATCAAGACCCGTATACGAATACCTCCGGCGAGATA	1112	
QY	1014	GIATGATATTCGCATCCATAIATGATACCGGACTTCGCCAGTAGTTATACAAAAATTTCT	1073	
DB	1113	CTAGCATACGCAATTCACGCGGATGTCAGCTCTCGTCCCTTACTTCTCGCGGTTCTT	1172	
QY	1074	GGCAAGGACTCTGTCTATGACGCTATTCGCGCTCAACATCAACTACACCCAGTCCATATA	1133	
DB	1173	GAATCGCCCATGGTGCAAAAGCAGCTTGGGTTCCCGGTGAAGTATACCATGTCGTGAGA	1232	
QY	1134	TGACGCTTACTACGCTTTCCAGCAAAACAGCGACTTTGCTGCCCCCAACTTCATCGAAGA	1193	
DB	1233	GGCAGTGGGGAACAGTTTCGCTCGACGGCGGATTTCCCGGAATGATCCCGGAATGATCCCGCGCGGAAT	1292	
QY	1194	CCTCGAGGAGA-----TCCITGCTTCCCGCTGGGTGTCTCCCTCATCTATGCGCA	1244	





Db 1172 GTGGGAGCCGCTG-----CGTTCAGGTCAAGCGCGCAACTTCAGTAGTGGG 1222  
QY 1331 GTACACGCC---CNGAAGTCACAGCGGTGAGTATGGGAAATCGCGAGTATGCTA 1387  
Db 1223 GTACACCAACATGTCACCTCGGATGAGTGCACACAGCGCCAGGTGCGCCAGCGGGC 1282  
QY 1388 AFTTCCTCTACTCGCTGTAGAGCGAGGCGCATGAAGTCCATACACGCCCATCG 1447  
Db 1283 AATTGCTCTTGTGCGAGTGTATGAGATGGACATGAGGTTCCTTCTATCACCTTGC 1342  
QY 1448 CTTCCCTGCAATGTTTAAACCGGACTATCTTCGTTGGGATATCGCAGAGGGCGAGAAGA 1507  
Db 1343 TTCCGCTGAGATGTTTGAAGCGGTCTATGCGCGCAAGGATGTGGCAGCGGAAGATTC 1402  
QY 1508 AGATCTGGCCAGCTACAGACGAATGGAAACGGCTA 1543  
Db 1403 CCATCTCGTCGAGTTTACAGACGCTGGCGACGCCCA 1438  
RESULT 8  
ABZ78184  
ID ABZ78184 standard; DNA; 2940 BP.  
XX AC ABZ78184;  
XX DT 24-APR-2003 (first entry)  
XX DE A. niger serine carboxypeptidase gene #1.  
XX KW Protease; fungal infection; aspergillosis; food; tanning; detergent;  
KW protein solubility; viscosity; taste; texture; nutritional value;  
KW gene; ds.  
XX OS Aspergillus niger.  
XX PN WO200268623-A2.  
XX PD 06-SEP-2002.  
XX PF 22-FEB-2002; 2002WO-EP01984.  
XX PR 23-FEB-2001; 2001EP-0200657.  
XX PR 23-FEB-2001; 2001EP-0200658.  
XX PR 23-FEB-2001; 2001EP-0200660.  
XX PR 26-FEB-2001; 2001EP-0200706.  
XX PR 26-FEB-2001; 2001EP-0200707.  
XX PR 26-FEB-2001; 2001EP-0200708.  
XX PR 26-FEB-2001; 2001EP-0200719.  
XX PR 28-MAR-2001; 2001EP-0000075.  
XX PR 28-MAR-2001; 2001EP-0000078.  
XX PR 28-MAR-2001; 2001EP-0000080.  
XX PR 28-MAR-2001; 2001EP-0000087.  
XX PR 28-MAR-2001; 2001EP-0000088.  
XX PR 21-MAY-2001; 2001EP-0000156.  
XX PR 21-MAY-2001; 2001EP-0000159.  
XX PR 21-MAY-2001; 2001EP-0000160.  
XX PR 21-MAY-2001; 2001EP-0000162.  
XX PR 21-MAY-2001; 2001EP-0000165.  
XX PR 21-MAY-2001; 2001EP-0000166.  
XX PR 21-MAY-2001; 2001EP-0000169.  
XX PR 21-JUN-2001; 2001EP-0000240.  
XX PR 21-JUN-2001; 2001EP-0000242.  
XX PR 21-JUN-2001; 2001EP-0000244.  
XX PR 21-JUN-2001; 2001EP-0000246.  
XX PR 12-JUL-2001; 2001EP-0000280.  
XX PR 12-JUL-2001; 2001EP-0000285.  
XX PR 30-JUL-2001; 2001EP-0000323.  
XX PR 30-JUL-2001; 2001EP-0000327.  
XX PR 02-AUG-2001; 2001EP-0000341.  
XX PR 02-AUG-2001; 2001EP-0000342.  
XX PR 02-AUG-2001; 2001EP-0000343.  
XX PR 02-AUG-2001; 2001EP-0000344.  
XX PR 09-AUG-2001; 2001EP-0000357.

PR 16-AUG-2001; 2001EP-0000374.  
PR 16-AUG-2001; 2001EP-0000377.  
PR 20-SEP-2001; 2001EP-0000478.  
PR 20-SEP-2001; 2001EP-0000483.  
PR 22-OCT-2001; 2001EP-0000552.  
PR 22-OCT-2001; 2001EP-0000553.  
PR 22-OCT-2001; 2001EP-0000554.  
PR 22-OCT-2001; 2001EP-0000556.  
PR 22-OCT-2001; 2001EP-0000557.  
PR 13-NOV-2001; 2001EP-0000454.  
PR 21-DEC-2001; 2001EP-00005117.  
XX (STAM ) DSM NV.  
XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;  
PI Klugbauer S, Wagner C, Britz A, Von Gustedt W, Heinrich O;  
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;  
PI Stiebler J, Albang R;  
XX WPI: 2002-723203/78.  
DR P-PSDB; ABR38817.  
DR Novel isolated protease polypeptide useful in laboratory, clinical,  
XX pharmaceutical, chemical, diagnostic, personal care and industrial  
XX applications  
XX Claim 1; Page 109-111; 394pp; English.  
XX The invention relates to a novel isolated protease polypeptide. A  
XX polypeptide or polynucleotide of the invention is useful for diagnosing a  
XX fungal infection such as aspergillosis, or as a query sequence to perform  
XX a search against public databases. A polypeptide of the invention is  
XX useful in a selected number of industrial or pharmaceutical processes, in  
XX laboratory or clinical processes, in food industry (baking, brewing,  
XX cheese manufacturing, meat tenderising), in tanning industry and in the  
XX manufacture of biological detergents. A polypeptide may also be useful  
XX for improving protein solubility, extraction yields, viscosity or taste,  
XX texture, nutritional value, minimising of antigenicity or  
XX anti-nutritional factors, colour or functionality as well as processing  
XX aspects like filterability of the proteinaceous raw material. The  
XX sequences shown in ABZ78180-ABZ78236 represent genes encoding the  
XX A. niger proteases of the invention.  
XX SQ Sequence 2940 BP; 691 A; 818 C; 732 G; 699 T; 0 other;  
Query Match 7.3%; Score 116.8; DB 24; Length 2940;  
Best Local Similarity 52.5%; Pred No. 5e-24;  
Matches 334; Conservative 0; Mismatches 287; Indels 15; Gaps 3;  
QY 914 ACCGCACCGCATTAGCTGACTACGCGCTCTGCGCGCAAGCCACCATGTGCGAGGACA 973  
Db 1492 ACTGCGCGCGCGAGGATCGACGAGATCTGCAGCACTGCGCAGCTTTTTCGCGCAAG 1551  
QY 974 ATGTTGAGGGCCATACACTACGCGCTTGTGTTGGTGTGTATGATATTCGCGATCCAT 1033  
Db 1552 AGGTGCAAAACGCTACGACATTTACTCGGTGGGATGAGTATGATCTTGTGACTCA 1611  
QY 1034 ATGATGACCGGACTCGCGCAAGTTATTACAAATAATTTCTGGCAAGGACTGTCTCATGG 1093  
Db 1612 CTCCCGACCGGTTCCTTACGAGTCTACGTTAGCTACCTGACAAAGCGTCGCGCAGG 1671  
QY 1094 AGGCTATCGGCGTCAACATCAACTACACCGTCCCAATATGACGTCTACTACGCTTCC 1153  
Db 1672 CGGCATCGGCGCATACATCAATATACAGGAGCAACACGCTTTGGACTCGCTTTT 1731  
QY 1154 AGCAACAGCGGACTTTGCTGGCCC---AACTTCATCGAAGACTCGAGGAGATCTTGTG 1210  
Db 1732 CGTCCACCGGTGACGACGCGGCGACTCATGAACACCATCCAGATGTGGCAGACTCTCA 1791  
QY 1211 CTCCTCCCGTGGGTGTTCTCCCTCATCTATGGGAGCGCGATTTACATCTCAACTGTTCG 1270  
Db 1792 AACAGGTGTCACGGTGGTTCATGTACGCGCGGGATCGCGACTATATACTGCAACTGCTGG 1851





QY 1286 CCTCGTCGCAACTACTCCCAAGCGGCCAGTTCGGAAGCGCAGGTACAGCCCTCGA 1345  
Db 1287 GCTTTCGAGTTCGATTCCTGATACCAATTTGCGGACAGCGGAGTACCTCCACTCC 2130  
QY 1346 AAGTCAACGGCTCGAGTAGTGGGAACATCGCGAGTATGGTAAATTCCTCTCCTACACGCGG 1405  
Db 1347 TTAACGCGCCGAGCGGATCAGCGCAATGACCGCCAGCTGGGCACTACAGCTTCACCTCGCG 2190  
QY 1406 TCTATGAGCGAGCGCATGAGTCCCACTATACGAGCCCATCGCTCCCTGCAATTCCTTA 1465  
Db 1407 TCTTCAAGCGCGGATGAGTCCCTCTACAGCTCTCGCGGCTATGAGATCTTCA 2250  
QY 1466 ACGGAGTATCTCGGTTCGGATATCGCAGAGGCGCAGAGAAGATCTGCGCCAGCTACA 1525  
Db 1467 TCGGGCGGACATCAACAAGATATCCCTACTGGCCTCTGCTGTGATGAGCAATTC 2310  
QY 1526 AGACGAATGGAAC 1538  
Db 2311 AGTCGGTTGGACC 2323  
RESULT 10  
ABZ78232  
ID ABZ78232 standard; DNA; 3221 BP.  
XX AC ABZ78232;  
XX AC ABZ78232;  
DT 24-APR-2003 (first entry)  
XX A. niger serine carboxypeptidase gene #4.  
DE Protease; fungal infection; aspergillosis; food; tanning; detergent;  
KW protein solubility; viscosity; taste; texture; nutritional value;  
KW gene; ds.  
XX Aspergillus niger.  
XX W0200268623-A2.  
PN 06-SEP-2002.  
PD 22-FEB-2002; 2002W0-EF01984.  
PF 23-FEB-2001; 2001EP-0200557.  
PF 23-FEB-2001; 2001EP-0200658.  
PF 23-FEB-2001; 2001EP-0200659.  
PF 26-FEB-2001; 2001EP-0200706.  
PF 26-FEB-2001; 2001EP-0200707.  
PF 26-FEB-2001; 2001EP-0200708.  
PF 26-FEB-2001; 2001EP-0200719.  
PF 28-MAR-2001; 2001EP-0000075.  
PF 28-MAR-2001; 2001EP-0000078.  
PF 28-MAR-2001; 2001EP-0000080.  
PF 28-MAR-2001; 2001EP-0000087.  
PF 28-MAR-2001; 2001EP-0000088.  
PF 21-MAY-2001; 2001EP-0000156.  
PF 21-MAY-2001; 2001EP-0000159.  
PF 21-MAY-2001; 2001EP-0000160.  
PF 21-MAY-2001; 2001EP-0000162.  
PF 21-MAY-2001; 2001EP-0000165.  
PF 21-MAY-2001; 2001EP-0000166.  
PF 21-MAY-2001; 2001EP-0000168.  
PF 21-JUN-2001; 2001EP-0000240.  
PF 21-JUN-2001; 2001EP-0000242.  
PF 21-JUN-2001; 2001EP-0000244.  
PF 21-JUN-2001; 2001EP-0000246.  
PF 12-JUL-2001; 2001EP-0000280.  
PF 12-JUL-2001; 2001EP-0000285.  
PF 30-JUL-2001; 2001EP-0000323.  
PF 30-JUL-2001; 2001EP-0000327.  
PF 02-AUG-2001; 2001EP-0000341.  
PF 02-AUG-2001; 2001EP-0000342.  
PF 02-AUG-2001; 2001EP-0000343.

02-AUG-2001; 2001EP-0000344.  
09-AUG-2001; 2001EP-0000357.  
PR 16-AUG-2001; 2001EP-0000374.  
PR 16-AUG-2001; 2001EP-0000377.  
PR 20-SEP-2001; 2001EP-0000478.  
PR 20-SEP-2001; 2001EP-0000483.  
PR 22-OCT-2001; 2001EP-0000552.  
PR 22-OCT-2001; 2001EP-0000553.  
PR 22-OCT-2001; 2001EP-0000554.  
PR 22-OCT-2001; 2001EP-0000556.  
PR 22-OCT-2001; 2001EP-0000557.  
PR 22-OCT-2001; 2001EP-0000558.  
PR 15-NOV-2001; 2001EP-0004464.  
PR 21-DEC-2001; 2001EP-0005117.  
XX (STAM ) DSM NV.  
XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel B;  
PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;  
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;  
PI Stiebler J, Albang R;  
XX  
DR WPI: 2002-723203/78.  
DR P-PSDB; ABR38865.  
XX Novel isolated protease polypeptide useful in laboratory, clinical,  
PT pharmaceutical, chemical, diagnostic, personal care and industrial  
PT applications  
XX  
PS Claim 1; Page 189-191; 394pp; English.  
XX The invention relates to a novel isolated protease polypeptide. A  
CC polypeptide or polynucleotide of the invention is useful for diagnosing a  
CC fungal infection such as aspergillosis, or as a query sequence to perform  
CC a search against public databases. A polypeptide of the invention is  
CC useful in a selected number of industrial or pharmaceutical processes, in  
CC laboratory or clinical processes, in food industry (baking, brewing, in  
CC cheese manufacturing, meat tenderising), in tanning industry and in the  
CC manufacture of biological detergents. A polypeptide may also be useful  
CC for improving protein solubility, extraction yields, viscosity or taste,  
CC texture, nutritional value, minimising of antigenicity or  
CC anti-nutritional factors, colour or functionality as well as processing  
CC aspects like filterability of the proteinaceous raw material. The  
CC sequences shown in ABZ78180-ABZ78236 represent genes encoding the  
CC A. niger proteases of the invention.  
XX  
SQ Sequence 3221 BP; 739 A; 773 C; 869 G; 840 T; 0 other;  
Query Match 6.7%; Score 107.8; DB 24; Length 3221;  
Best Local Similarity 48.5%; Pred. No. 2.6e-21;  
Matches 449; Conservative 0; Mismatches 437; Indels 39; Gaps 4;  
QY 614 TCAGTCTATGGACGGAGAGCTATGGAGGGGCACTATGGTCCGTGCAATTCATCATTTT 673  
Db 1522 TGAGTATATGACGGAGTCATATGGGGAGCGATGACCGGCATACACGGGCTTTTC 1581  
QY 674 ACGAGCAAGTACAGAAATGCCAACGGTAGTGTAAAT---GGTCTTCAGCTTATTTCA 730  
Db 1582 AGGAGATCAATGAGAGGATTCAGAGTGGGGAGGTAAAGCAGCGGAGAGATCCATTGG 1641  
QY 731 ACTCTCTGGGATTTATTAACGGCATCATCGAGAGCGGATCCAGCCCTTACTACCTGTG 790  
Db 1642 ATACGCTGGGCATTTATCAATGGGTGTGGATTTACTGTCGAGGTCCCTTCGTCCTG 1701  
QY 791 AATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCACGAGACCGCTCTACAACTACA 850  
Db 1702 ACCAGCGTATTAACATACATACGATGATGAGGAAATCAATCGCAGCTCTACGACGSG 1761  
QY 851 TGAAGTTTGCCAAACCAATGCCAATGGTTCAGAGATTTGATTTCCACCTG----- 902  
Db 1762 CTAAGGATAGTTGGAGCAAGCCTGGCGGTGAGGAGATATGATCATCGAGTGTGCGGATG 1821  
QY 903 -----CAAACAGACAACCGCACCGCAATTAGTGTACTAGCGCCTCTGCG 946

Db 1822 CTGGCGAGCTGGAGATCCCTCATGTATGGCGACAATGAGCGGTAAATAGACTCTCG 1881  
 QY 947 CCGAAGCCACCAACATGTCAGGGACAATGTTGAGGGCCATACACGCCCTTTCGIGTC 1006  
 Db 1892 AGGAGCGCTCGGACTACTGTTGCGGGAGATCAAGAGCTGTATAGCAATACCTCCGGC 1941  
 QY 1007 GTGGTGTATGATATTCGGCATCCATATGATGACCCGACTCCGCAAGTTATACAACA 1066  
 Db 1942 GAGGATACAGACATAGCGCATTTACGCGCGATGACGCTCTGTCCTTACTTCGTCG 2001  
 QY 1067 AATTCGCGAAGAGCTCTGTATGAGCGTATCGGCGTCAACATCACTACACCCAST 1126  
 Db 2002 GGTCTTGAATCGCCATGTTGTCGCAAGGCACTTGGGTCCTCCGCTGAACATACCATGT 2061  
 QY 1127 CCAATAATGAGCTCTACTACCTTCCAGCAACAGCGGACT-----TTGCTGGC 1177  
 Db 2062 CGTCAGAGGCAAGTGGGCAACAGTTTCGCTCGACGGCGATATCCGGGAAATGATCCCC 2121  
 QY 1178 CCACTTCATCGAAGACTCGAGAGATTCCTTCTCTCCCGTGGTGTCTCCCTCATCT 1237  
 Db 2122 GCGAATGATCGGGATATGATACCTGCTTGTACTCCGCTGTCAGGTCGCTATGTAT 2181  
 QY 1238 ATGCGGACCGGATATGATGCACTGCTGTCGCGGTGAGCGGCTTCCCTCGCTCGGA 1297  
 Db 2182 ATGCGGACCGGATGCTTGTCCGTCGCGCGCGGGAAGATGTCAGCGTCTGCTG 2241  
 QY 1298 ACTACTCCCAAGCGCCAGTTCGGAAGCGCAGGCTACACGCCCTCGAAAGTCAACGGG 1357  
 Db 2242 AGTACGAGGATCGGAGAAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2301  
 QY 1358 TCGAGTATGGG---AACTCGCGAGTATGATGATATTTCTCTTCACTCGGCTCTATGAG 1414  
 Db 2302 CCTACGTTGGGGTCTAGTAAGCAGTATGGGAACCTCTGTTCAACGCGTCTTTCAGG 2361  
 QY 1415 CAGCGCATGAAGTCCCATACACGCCATCCGCTCCCTCGAATGTTTAAACGGGACTA 1474  
 Db 2362 CGGCGCATGAGTGCCTATTTATCAGCCCGAAGCGGTATGAGATTTTATTCGCGCTC 2421  
 QY 1475 TCTTCGTTGGGATATCGCAGAGG 1499  
 Db 2422 AGTTTAATGGGATATTCGACGGG 2446

## RESULT 11

AAFL12116  
 ID AAFL12116 standard; cDNA; 678 BP.  
 AC AAF12116;  
 XX  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Aspergillus oryzae EST SEQ ID NO:4639.  
 XX  
 KW Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 OS Aspergillus oryzae.  
 XX  
 PN WO200056762-A2.  
 XX  
 XX 28-SEP-2000.  
 PD  
 XX  
 PF 22-MAR-2000; 2000WO-US07781.  
 XX  
 XX 22-MAR-1999; 99US-0273623.  
 PR  
 XX (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO NORDISK AS.  
 XX

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 XX WPI: 2000-594572/56.  
 DR  
 XX  
 PT Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags -  
 XX  
 PS Claim 88; Page 1971; 3161pp; English.  
 XX

The present invention describes a method for monitoring differential expression of genes in a first filamentous fungal (FF) cell relative to expression of the same genes in one or more second filamentous fungal cells. The method uses fluorescence-labeled nucleic acids isolated from the FF cells and a substrate of expressed sequence tags (EST). The ESTs are used in the methods for monitoring differential expression of genes in a first filamentous fungal (FF) cell relative to expression of the same genes in one or more second filamentous fungal cells. Monitoring the global expression of genes from FF cells allows the production potential of the microorganisms to be improved. New genes may be discovered, possible functions of unknown open reading frames can be identified and gene copy number variation and stability can be monitored. The expression of genes can be used to study how FF cells adapt to changes in culture conditions, environmental stress, spore morphogenesis, recombination, metabolic or catabolic pathway engineering. Using ESTs provides several advantages over genomic or random cDNA clones including elimination of redundancy as one spot on an array equals one gene or open reading frame, and organisation of the microarrays based on function of the gene products to facilitate analysis of the results. AAF07478 to AAF11247 represents ESTs from *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are all specifically claimed in the present invention.

Sequence 678 BP; 153 A; 188 C; 175 G; 161 T; 1 other;

Query Match 6.48; Score 103; DB 21; Length 678;  
 Best Local Similarity 53.0%; Pred. No. 3,2e-20;  
 Matches 273; Conservative 0; Mismatches 230; Indels 12; Gaps 2;  
 QY 1001 CTGCTGCTGGTGTATGATATTCGCGCATCATATGATGACCGGCTCCGCGAGTTATT 1060  
 Db 3 CTGCGAGAGCTATTACTATCTGCGATTTTACCCCGGATCTTTCCCGCTCCGTTACT 62  
 QY 1061 ACAACAAATTCGGCAAGGACTCTGTATGACGCTATCGGCTCAACATCAACTACA 1120  
 Db 63 ACTTTGTTTCTTAAACACGACTGGGTCAGAGAGCCCTCGGGTTCGCGTTAATTTCA 122  
 QY 1121 CCCAGTCCATATAGCTCTACTAGCTTTTCCAGCAACAGCGGCTTTTCTGCGCCA 1180  
 Db 123 CGGAATCTGTGTATAGCTCTACATATGGCTTTTCCGCAACAGGTAACCCAGCTCCG 182  
 QY 1181 ACTTCATCGAAGACCTCGAGGAGATC-----CTTGTCTCTCCCGTGGTGTCTCC 1231  
 Db 183 ATGTACCGGTTACCTGGAGGATATCGCATGCTGTGCTTGTACTCTGCGATCAAAAGTTGCC 242  
 QY 1232 TCATCTATGCGGACCGCGATTACATCTGCAACTGTTCGCGCTCAGCGGTTTCCTCTCG 1291  
 Db 243 TAGTGTATGCGATCGGATATGCGATGCCCCCTGGACGAGGAGAGAGTGAAGTGA 302  
 QY 1292 CTGCGCACTACTCCAGCGCCGCTTCGAGGCGAGGTTACAGCCCTTCAAGTCA 1351  
 Db 303 AGGTGAGTACTCGGATCGGCAAGTTCGCTCTGCGGTTACGCCCCCTCTGAGACCA 362  
 QY 1352 ACGGCTCGAGTATGGGAACT---CGGAGTATGTAATTTCTCTTCACTCGGCT 1408  
 Db 363 ATGCTCATATCTAGTGTGCTGGTCCGACAGTACGGAACATCTCGTTCCTCGGTT 422  
 QY 1409 ATGAGCGAGGCCATGAAGTCCCATACTACACGCCATCCCTCCCTGCAATGTTTAC 1468  
 Db 423 TCGAAGCGGTCATGAGTGGCGGATATCACACCTGAAACGGGTATGATGATCTTCCACC 482

QY 1469 GGACTATCTCGTTGGATATCGCAGAGGGCCAG 1503  
DB 483 GAGCATTTGTTCAACAGACATTCGACGGGCAAG 517

RESULT 12

AAFL13804  
ID AAF13804 standard; cDNA; 636 BP.

AC AAF13804;

DT 13-MAR-2001 (first entry)

DE Aspergillus oryzae EST SEQ ID NO:6327.

Multiple gene expression; filamentous fungal cell; EST;  
expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
culture condition; environmental stress; spore morphogenesis;  
metabolic pathway engineering; catabolic pathway engineering; ss.

OS Aspergillus oryzae.

WO2000056762-A2.

28-SEP-2000.

22-MAR-2000; 2000WO-US07781.

22-MAR-1999; 99US-0273623.

(NOVO ) NOVO NORDISK BIOTECH INC.  
(NOVO ) NOVO NORDISK AS.

Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

WPI; 2000-594572/56.

Monitoring differential expression of genes in filamentous fungal cells  
uses fluorescence-labeled nucleic acids isolated from the cells and a  
substrate of expressed sequence tags -

Claim 88; Page 2600; 3161pp; English.

The present invention describes a method for monitoring differential  
expression of genes in a first filamentous fungal (FF) cell relative to  
expression of the same genes in one or more second filamentous fungal  
cells. The method uses fluorescence-labeled nucleic acids isolated from  
the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
are used in the methods for monitoring differential expression of genes  
in a first filamentous fungal (FF) cell relative to expression of the  
same genes in one or more second filamentous fungal cells. Monitoring  
the global expression of genes from FF cells allows the production  
potential of the microorganisms to be improved. New genes may be  
discovered, possible functions of unknown open reading frames can be  
identified and gene copy number variation and stability can be  
monitored. The expression of genes can be used to study how FF cells  
adapt to changes in culture conditions, environmental stress, spore  
morphogenesis, recombination, metabolic or catabolic pathway  
engineering. Using ESTs provides several advantages over genomic or  
random cDNA clones including elimination of redundancy as one spot on an  
array equals one gene or open reading frame, and organisation of the  
microarrays based on function of the gene products to facilitate  
analysis of the results. AAF07478 to AAF11247 represents ESTs from  
Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
niger; AAF11854 to AAF14878 represents ESTs from Trichoderma reesei; and  
AAFL14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
all specifically claimed in the present invention.

Sequence 636 BP; 148 A; 177 C; 155 G; 156 T; 0 other;

Query Match 5.5%; Score 87.8; DB 21; Length 636;  
Best Local Similarity 54.9%; Pred. No. 1.1e-15;

Matches 195; Conservative 0; Mismatches 157; Indels 3; Gaps 1;  
QY 120 GGGCGTCTCGGACATPACCCGGGTCTCAAAATCTACTCTGTGATATGTGAC---ACCTC 176  
DB 79 GAGCGTCTCGGTATCGACCCGAATGTGAAGCAATACACTGTGTATCTGACGATAACGG 138  
QY 177 TCCGAGTCCCATACCTTCTTCTGGTCTTCGAAAGCCAGACATAAACCCAGAACTGCACC 236  
DB 139 TAAAGACAAGCATCTGTCTACTGTCTTCGAAATCTCGCAATGATGCTTAAGAACGCC 198  
QY 237 TATCACATTTGTGTTGAATGGTGGCCCTGGAAGCATTTCTTGTGATGGTCTCTTCGAAGA 296  
DB 199 CGTCGTCTGTGCTGAAAGTGGCCCGGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 258  
QY 297 GTTGGGCCCTTGGCCATGTCAATTCGACTTTTGAATGATACATCAACCTCAGCTGGGAA 356  
DB 259 GCTGGGACCTTAGCAGCATGAGAACATTAAGCCCGTCTACAATGACTTCTCGTSGAA 318  
QY 357 CGAGGTCTCCAATTACTATTCTGTCGCCAGCCATTCGGAGTTCGGCTTTCATATAGTA 416  
DB 319 CTCTAAGCGCTCGTCAATTTCTTGTATCAGCCCGTGAACGTGGCTATTCCTACATGG 378  
QY 417 TACGTTGATGGTCCATTAAACCTGTAACTGGGTGCTGCAAAATTCGAGCTTT 471  
DB 379 CTGCGTGTAGGACACTGTCTGCTGGCAGGATGTCTACGCTCTGCTCT 433

RESULT 13

AAQ55347  
ID AAQ55347 standard; cDNA; 2503 BP.

AC AAQ55347;

25-MAR-2003 (updated)  
20-JUL-1994 (first entry)

Sequence of gene Kl.PRC1 encoding protease C.

Protease; yeast; proteolysis; ss.

Kluyveromyces lactis.

Key Location/Qualifiers  
CDS 387..1862  
/\*tag= a

WO9400579-A1.

06-JAN-1994.

23-JUN-1993; 93WO-FR00623.

25-JUN-1992; 92FR-0007785.

(RHON ) RHONE POULENC ROBER SA.

Fleer R, Fournier A, Yeh P;

WPI; 1994-026215/03.

P-PSDB; AAR48059.

New Kluyveromyces yeast with modified protease gene - esp. used  
for high yield prodn. of recombinant protein, also DNA encoding  
yeast protease and derived peptide(s)

Claim 25; Page 28-31; 49pp; English.

The protease gene is to be modified in order to render it (partially)  
incapable of producing the natural protein; or result in a non-  
functional protease or in a protease with modified proteolytic  
activity. The modifications can be introduced in vitro or in situ by  
standard genetic engineering techniques or by exposure to mutagenic  
agents.



PR 21-MAY-2001; 2001EP-0000156.  
PR 21-MAY-2001; 2001EP-0000159.  
PR 21-MAY-2001; 2001EP-0000160.  
PR 21-MAY-2001; 2001EP-0000162.  
PR 21-MAY-2001; 2001EP-0000165.  
PR 21-MAY-2001; 2001EP-0000166.  
PR 21-MAY-2001; 2001EP-0000168.  
PR 21-JUN-2001; 2001EP-0000240.  
PR 21-JUN-2001; 2001EP-0000242.  
PR 21-JUN-2001; 2001EP-0000244.  
PR 21-JUN-2001; 2001EP-0000246.  
PR 12-JUL-2001; 2001EP-0000280.  
PR 12-JUL-2001; 2001EP-0000285.  
PR 30-JUL-2001; 2001EP-0000323.  
PR 30-JUL-2001; 2001EP-0000327.  
PR 02-AUG-2001; 2001EP-0000341.  
PR 02-AUG-2001; 2001EP-0000342.  
PR 02-AUG-2001; 2001EP-0000343.  
PR 02-AUG-2001; 2001EP-0000344.  
PR 09-AUG-2001; 2001EP-0000357.  
PR 16-AUG-2001; 2001EP-0000374.  
PR 16-AUG-2001; 2001EP-0000377.  
PR 20-SEP-2001; 2001EP-0000478.  
PR 20-SEP-2001; 2001EP-0000483.  
PR 22-OCT-2001; 2001EP-0000552.  
PR 22-OCT-2001; 2001EP-0000553.  
PR 22-OCT-2001; 2001EP-0000554.  
PR 22-OCT-2001; 2001EP-0000556.  
PR 22-OCT-2001; 2001EP-0000557.  
PR 22-OCT-2001; 2001EP-0000558.  
PR 15-NOV-2001; 2001EP-0004464.  
PR 21-DEC-2001; 2001EP-0005117.  
XX  
FA (STAM ) DSM NV.

XX

XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;  
PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;  
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;  
PI Stiebler J, Albang R;

XX WPI; 2002-723203/78.

DR P-PSDB; ABR38819.

XX Novel isolated protease polypeptide useful in laboratory, clinical,  
PT pharmaceutical, chemical, diagnostic, personal care and industrial  
PT applications

XX Claim 1; Page 208; 394pp; English.

XX The invention relates to a novel isolated protease polypeptide. A  
CC polypeptide or polynucleotide of the invention is useful for diagnosing a  
CC fungal infection such as aspergillosis, or as a query sequence to perform  
CC a search against public databases. A polypeptide of the invention is  
CC useful in a selected number of industrial or pharmaceutical processes, in  
CC laboratory or clinical processes, in food industry (baking, brewing,  
CC cheese manufacturing, meat tenderising), in tanning industry and in  
CC manufacture of biological detergents. A polypeptide may also be useful  
CC for improving protein solubility, extraction yields, viscosity or taste,  
CC texture, nutritional value, minimising of antigenicity or  
CC anti-nutritional factors, colour or functionality as well as processing  
CC aspects like filterability of the proteinaceous raw material. The  
CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the  
CC A. niger proteases of the invention.

XX Sequence 1611 BP; 326 A; 484 C; 422 G; 379 T; 0 other;

Query Match 5.1%; Score 82.4; DB 24; Length 1611;  
Best Local Similarity 55.2%; Pred. No. 7.7e-14;  
Matches 160; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 169 GACACCTCCCGAGTCCATACCTCTCTGGTCTTCCGAAGCCAGACATAACCCAGAA 228

|||||  
DQ 415 GACAACGAGAAATGATAGCATTTGTTACTGGTCTTCGAGTCTCGCAATGACCCCGAG 474

QY 229 ACTGCACCTATACATTTGTTGAATGGTGGCCCTGGAAAGCGATTCTTTGATCGGTCTC 288  
|||  
Db 475 AATGATCCCGTGTGTTCTGTGGCTGACACGGTGGCCCTGGTGTCTTCCCTCACGGTCTC 534  
|||  
QY 289 TTCGAGAGATTGGGCCCTTCCCATGTCAATTGCGACTTTTGTGATGACTACATCAACCCCTCAC 348  
|||  
Db 535 TTCAATGGAGCTTGGCCCTAGCAGCATCAACAAGAAGATCCAGCCGGTCTACATGACTAC 594  
|||  
QY 349 TCGTGAACGAGGTCTCCAAATTTACTATTCTCTGCCAGCAATGGGAGTGGGCTTTTCA 408  
|||  
Db 595 GCTTGGAACTCCCAACGGCTCGGTGATCTTCTTGACACAGCTGTCAATGTCGGTTACTCC 654  
|||  
QY 409 TATAGTGATACGGTTGATGGTCCATTAAACCCCTGTAACCTGGGGTCTGTCGA 458  
|||  
Db 655 TACAGTAACTCTGCTGTGTCAGCGACACGTCGTCGTGGCAAGGAGCTCTA 704  
|||

Search completed: September 16, 2003, 18:14:13  
Job time : 559 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 16, 2003, 15:55:43 : Search time 72.5 Seconds  
(without alignments)  
7040.892 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662  
Perfect score: 2904  
Sequence: 1 cttccaggaggtacaccggc.....gcattgccagtgttgatg 1508

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-O=/cn2.1/USPTO.spool\_p/US09712338/runat\_16092003\_144324\_14334/app\_query.fasta\_1.1799  
-DB=A\_Geneseq\_19Jun03 -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09712338.ecgn\_1.1.92.0runat\_16092003\_144324\_14334 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THRAADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_19Jun03.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2878	99.1	554	19	AAW56099	A. oryzae ATCC2038
2	2338	80.5	551	23	ABR38864	A. niger serine ca
3	980.5	33.8	623	23	ABR38865	A. niger serine ca
4	946	32.6	554	23	ABR38859	A. niger serine ca
5	914.5	31.5	526	23	ABR38817	A. niger serine ca
6	636	21.9	508	23	ABG30049	S. cerevisiae BAX-
7	630	21.7	536	23	ABR38819	A. niger carboxype
8	630	21.7	556	17	AA96737	A. niger Bo-1 carb
9	620	21.3	557	17	AA96738	A. niger SPAG 2 ca
10	568.5	19.6	491	15	AA48059	Sequence of protea
11	541.5	18.6	550	23	ABG33281	C. albicans BAX-as
12	414	14.3	481	23	ABR38845	A. niger carboxype
13	383	13.2	482	21	AA30085	Arabidopsis thalia
14	383	13.2	502	21	AA30064	Arabidopsis thalia
15	366.5	12.6	455	23	ABR38843	A. niger carboxype
16	361.5	12.4	479	21	AA323905	Arabidopsis thalia
17	361.5	12.4	486	21	AA323904	Arabidopsis thalia
18	361	12.4	476	19	AAW72966	Human serine carbo
19	361	12.4	476	22	AA888381	Human membrane or
20	359	12.4	476	20	AAV13372	Amino acid sequenc
21	359	12.4	476	21	AA801407	Human TANGO 176.
22	359	12.4	476	22	AAU29228	Human PRO polypept
23	359	12.4	476	22	AA888587	Human hydrophobic
24	359	12.4	476	22	AA880240	Human PRO polypept
25	359	12.4	476	24	ABU69650	Human secreted/pro
26	359	12.4	476	24	ABU71316	Human secreted/pro
27	359	12.4	476	24	ABU71473	Human PRO polypept
28	359	12.4	476	24	ABU71919	Human secreted/pro
29	359	12.4	476	24	ABU65773	Human secreted/pro
30	359	12.4	476	24	ABU66106	Human secreted/pro
31	359	12.4	476	24	ABU67373	Human secreted/pro
32	359	12.4	476	24	ABU67610	Human secreted/pro
33	359	12.4	476	24	ABU64527	Human secreted/pro
34	359	12.4	476	24	ABU65468	Human secreted/pro
35	359	12.4	476	24	ABU58604	Human PRO polypept
36	359	12.4	476	24	ABU56140	Human secreted/pro
37	359	12.4	476	24	ABU57135	Human PRO polypept
38	359	12.4	476	24	ABU54375	Human secreted/pro
39	359	12.4	476	24	ABU10714	Human secreted/pro
40	359	12.4	477	22	AA25810	Human protein sequ
41	357.5	12.3	479	21	AA47178	Arabidopsis thalia
42	357.5	12.3	486	21	AA47177	Arabidopsis thalia
43	357	12.3	476	20	AA28570	Secreted peptide c
44	357	12.3	476	22	AAU39043	Human secreted pro
45	357	12.3	476	23	ABW55752	Human polypeptide

ALIGNMENTS

RESULT 1  
AAW56099  
ID AAW56099 standard; Protein; 554 AA.  
XX  
AC AAW56099;  
XX  
DT 27-AUG-1998 (first entry)  
XX  
DE A. oryzae ATCC20386 carboxypeptidase I protein.  
XX  
KW Carboxypeptidase I; flavour improving agent; hydrolysate; proteinaceous;  
KW food industry.  
XX  
OS Aspergillus oryzae.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..18  
FT /label= signal

Protein 19...554  
/label= carboxypeptidase I

102b

09-APR-1998

03-OCT-1997; 97WO-US17977.

27-NOV-1996; 96US-0757534.

04-OCT-1996; 96US-0726880.

(NOVO ) NOVO NORDISK BIOTECH INC.

(NOVO ) NOVO-NORDISK AS.

Berka R, Blinkovsky A, Brown K, Dammann C, Golightly E;  
Klotz A, Mathisen TE, Rey M;

WPI; 1998-240098/21.

N-PSDB; AAV28620.

Carboxypeptidase from Aspergillus oryzae - hydrolyses proteinaceous  
substrates, useful for improving flavour of foods

Claim 1; Fig 3; 82pp; English.

This sequence represents carboxypeptidase I from Aspergillus oryzae.  
This polypeptide has an optimal activity in the range of pH 3.0-7.5 at  
25 deg. C, optimal activity in the range of 55-60 deg. C at pH 4, and a  
residual activity of at least 65.5% after 30 minutes at pH 4.0 and  
60 deg. C. It also has the capacity to hydrolyse X from N-CBZ-Ala-X where  
N-CBZ is N-carboxy and X is any amino acid. The carboxypeptidases  
can be used for obtaining hydrolysates (which can be enriched in free  
glutamic acid or peptide bound glutamic acid residues) from proteinaceous  
substrates. The carboxypeptidases can be used in flavour-improving  
compositions in the food industry. The products can also be used for the  
production of polypeptides free of carboxypeptidase activity.

Sequence 554 AA;

Alignment Scores:  
Pred. No.: 1.43e-253 Length: 554  
Score: 2878.00 Matches: 536  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.10% Indels: 0  
DB: 19 Gaps: 0

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x AAW56099 (1-554)

QY 1 CTTCCAGGAAGTACACCGGCTCGCTCGGTAGAGACAGCTACCCAGAACCCACCGGG 60  
Db 19 LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly 38  
QY 61 GTCAAGACTCTTACACCGCAACAAATGTACCATCCGGTACAGGAACCCGGGCGAG 120  
Db 39 VallysThrLeuThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGlu 58  
QY 121 GCGGTCTGGAGACTACCCGGGTCTCAATCTTACTCTGATATGTCGACCTCTCC 180  
Db 59 GlyValCysGluThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 78  
QY 181 GAGTCCCATACCTTCTTGTGTTCTTCAAGCCAGACATACCCAGAACTGCACCTATC 240  
Db 79 GluSerHisThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProIle 98  
QY 241 ACATTGTGGTGAATGGTGGCCCTGGAGCGATTCTTTGATCGGTCTCTTCGAGAGTIG 300  
Db 99 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeu 118  
QY 301 GGCCCTTCCCATGCAATTCGACTTTTGTACGTACATCAACCTCTACTCTGTGAACGAG 360  
Db 119 GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerIrpAsnGlu 138

QY 361 GTCTCCAATTACTATTCTCTGTCCTCCAGCCATGGGAGTCGGCTTTTCATAGTAGTACG 420  
Db ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 158  
QY 421 GTTGATGGTCCATTAACCTGTAACTGGGTGCTGCGAAATTCAGCTTTTCAGAGATT 480  
Db ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal 178  
QY 481 CAGGGCGGTACCCAAACCATTTGATCCACTCTGATCCATACTACTACCAATCTTGGCCGAG 540  
Db GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGlu 138  
QY 541 GCGGTGGGAGATCCTGCAAGGATTCCTTAGTGAGTACTACCTAGCTTGGACTCTAGGGTG 600  
Db AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal 218  
QY 601 CAGCTAAAGGACTTCAGTCTATGACGGAGAGAGTATGGAGGCGACTATGTCCTGCATTC 660  
Db GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 238  
QY 661 TTCAATCATTTTACGAGCAGAATGAGAGAAATGCCAACGGTAGTGTATGTGTGTTCAG 720  
Db PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 258  
QY 721 CTTAATTTCAACTCTCTGGGAATTTAAGCGCATCATCGACGAGCGGATCCAGGCCCT 780  
Db LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleAspGluAlaIleGlnAlaPro 278  
QY 781 TACTACCTCAATTCGCTGTGCAACATACCTACGGTATCAAGGCTGTCAACGAGCAGCGTC 840  
Db TyrTyrProGluPheAlaValAsnThrTyrGlyIleLysAlaValAsnGluThrVal 298  
QY 841 TACAACCTACATGAAGTTTGGCAACCAATGCCAATGGTTGCCAGGATTTGATTTCCACC 900  
Db TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr 318  
QY 901 TGCAACAGACAAACCCGCGCATTAGCTAGCTACGCCCTCTCGCGGAGCGCACCAAC 960  
Db CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn 338  
QY 961 ATGTGCGGAGACATGTTAGGGGCCCATACTACGCCCTTTGCTGCTGCTGTGTATGAT 1020  
Db MetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgValTyrAsp 358  
QY 1021 ATTCGGCATCATATGATGACCCGACTCCGCCAGTATATACAAATTTCTTGGCAAG 1080  
Db IleArgHisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys 378  
QY 1081 GACTCTGTCATGGAGCTATCGGGTCAACATCAACTACACCCAGTCCCAATATGACGTC 1140  
Db AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspVal 398  
QY 1141 TACTACGCTTCCAGCAACAGCGACTTTGCTGGCCCAACTTCATCGAAGACCTCGAG 1200  
Db TyrTyrAlaPheGlnGlnThrGlyAspPheValIrpProAsnPheIleGluAspLeuGlu 418  
QY 1201 GAGATCCTTGTCTCCCGCTCGCTCTCCCTCATCTATGCGGCGCGGATTCATCTGC 1260  
Db GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys 438  
QY 1261 AACTGGTTCGGCGGTACGCGGTTTCCCTCGCTCGGAATCTCTCCAGCGCGCCAGTTC 1320  
Db AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPhe 458  
QY 1321 CGAAGCGCGGTACACGCCCTCGTAAAGTCAACGCGCTCGAGTAGTGGGGAACCTCGCAG 1380  
Db ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu 478  
QY 1381 TATGTAATTTCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTC 1440  
Db TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln 498





QY	361	GTCTCCAACTTACTATCTCTCCAGCATTGGAGTCGGCTTTCATATAGTGATACG	420		Db	499	ProileAlaAlaLeuGlnLeuPheAsnArgThrLeuPheGlyTrpAspIleAlaAlaGly	518	
Db	139	ValThrAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerGluThr	158		QY	1501	CAGAAAGAACTTCGGCCAGCTACAAGACGAATGAAGGCTACAGCTAGGCATACACAG	1560	
QY	421	GTGTATGGTCCATTAAACCTTAACCTGAGGGTCGGTGAATAATTCAGCTTTCGAGAGTT	480		Db	519	ThrThrGlnIleTrpProGluTyrSerThrAsnGlyThrSerGlnAlaThrHisThrGlu	538	
Db	159	GluAlaGlySerLeuAsnProPheThrGlyAlaValGluAsnAlaSerPheAlaGlyVal	178		QY	1561	TCGTCCGTGCGCTGCTACGGCTACGCTACGAGC	1590	
QY	481	CAGGGCCGTACCCACCAATGATGACCTCTGATGCAATACCTACCAATCTTCCCGCAGAG	540		Db	539	SerPheValProLeuSerThrAlaSerSer	548	
Db	179	GlnGlyArgTyrProValIleAspAlaThrIleAspThrThrAspIleAlaAlaArg	198		RESULT 3				
QY	541	GGCGCTGGAGATCTCGAAGATTCCTTAGTGACCTACCTAGCTTGGACTCTAGGGTG	600		ABR38865				
Db	199	AlaThrTrpGluValLeuGlnGlyPheLeuSerGlyLeuSerGlnLeuAspSerGluVal	218		ID	ABR38865	standard; Protein; 523 AA.		
QY	601	CAGTCTAGGACTTCACTATGAGCGGAGAGCATGAGGGGACATGATGCTCTGCTGATTC	660		XX	AC	ABR38865;		
Db	219	LysSerLysGluPheAsnLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe	238		XX	DT	24-APR-2003 (first entry)		
QY	661	TTCAATCAATTTTACGAGCAGATGAGAGNATTCGCAACGGTAGTGTATGCTGTTTCAG	720		XX	DE	A. niger serine carboxypeptidase polypeptide #4.		
Db	239	PheAsnHisPheTyrGluGlnAsnSerLysIleAlaSerGlyGluValAsnGlyValGln	258		XX	KW	Protease; fungal infection; aspergillosis; food; tanning; detergent;		
QY	721	CTTAATTTCAACTCTCTGGGAATTAATTAACGGCATCATCGACGAGCGATCCAGGCCCT	780		XX	KW	protein solubility; viscosity; taste; texture; nutritional value;		
Db	259	LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspAlaIleGlnAlaAsp	278		XX	KW	EC3.4.16.6.		
QY	781	TACTACCTCGAATTCGTGTGAACAATACCTACGGTATCAAGCTGTCAACGAGACCGTC	840		XX	OS	Aspergillus niger.		
Db	279	TyrTyrAlaAspPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnAspThrVal	298		XX	OS	W0200268623-A2.		
QY	841	TACAACCTACATGAGTTGTCACCAACCAATCCCAATGTTGCCAGGATTTGATTCACC	900		XX	PD	06-SEP-2002.		
Db	299	TyrAsnTyrMetLysPheAlaAsnThrMetProAsnGlyCysGlnAspGlnValAlaSer	318		XX	PF	22-FEB-2002; 2002WO-EP01984.		
QY	901	TGCAACAGACAAACCCAGCATTAGCTACCTACCTACCGCTCTCGCGCGAAGCCACCAAC	960		XX	PR	23-FEB-2001; 2001EP-0200657.		
Db	319	CysLysLeuThrAsnArgThrSerLeuSerAspTyrAlaIleCysThrGluAlaAlaAsn	338		XX	PR	23-FEB-2001; 2001EP-0200660.		
QY	961	ATGTGACGAGCAATGTTAGGGGCCAATACACCCCTTTCGTGGTGGTGTATGAT	1020		XX	PR	26-FEB-2001; 2001EP-0200706.		
Db	339	MetCysArgAspAsnValGluGlyProTyrTyrGlnPheGlyGlyArgGlyValTyrAsp	358		XX	PR	26-FEB-2001; 2001EP-0200707.		
QY	1021	ATTGCGCATCATATGATACCCGACCTCCGCAAGTTATTAACAATTTCTCGCAAG	1080		XX	PR	26-FEB-2001; 2001EP-0200708.		
Db	359	IleArgHisProTyrAsnAspProThrProProSerTyrPheValAspTyrLeuLys	378		XX	PR	26-FEB-2001; 2001EP-0200719.		
QY	1081	CACCTCTCATGACGCTATGCGCTCAACATCACTACACCCAGTCCCAATATGACGTC	1140		XX	PR	28-MAR-2001; 2001EP-0000075.		
Db	379	AspSerValMetAspAlaIleGlyValAspIleAsnTyrThrGluSerSerGlyGluVal	398		XX	PR	28-MAR-2001; 2001EP-0000078.		
QY	1141	TACTACGCTTTCAGCAACAGCGCATTGTCTGCGCCCACTTCATCGAAGACTCGAG	1200		XX	PR	28-MAR-2001; 2001EP-0000080.		
Db	399	TyrTyrAlaPheGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu	418		XX	PR	28-MAR-2001; 2001EP-0000087.		
QY	1201	GAGATCTCTGCTCCCGCTCGGTCTCCCTCATCTATGCGCGCGCGATTCATCTGTC	1260		XX	PR	28-MAR-2001; 2001EP-0000088.		
Db	419	GluIleLeuGlnLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys	438		XX	PR	21-MAY-2001; 2001EP-0000156.		
QY	1261	AACTGGTTCGGCGGTAGCGGCTTTCCTCTCGCTGCACTACTCCAGCGCGCCAGTTC	1320		XX	PR	21-MAY-2001; 2001EP-0000159.		
Db	439	AsnTrpPheGlyGlyGlnAlaIleSerLeuAlaValAsnTyrProHisAlaAlaGlnPhe	458		XX	PR	21-MAY-2001; 2001EP-0000160.		
QY	1321	CGAAGCGGAGGTACAGCCCTGAAAGTCAACGGCTCGAGTATGGGAAACTCGCGAG	1380		XX	PR	21-MAY-2001; 2001EP-0000162.		
Db	459	ArgAlaAlaGlyTyrThrProMetThrValAspGlyValGluTyrGlyLuthrArgGlu	478		XX	PR	21-MAY-2001; 2001EP-0000165.		
QY	1381	TATGTAATTTCTCTTCACTCGGCTCTATGAGCGAGCCCATGAGTCCCATCTACACAG	1440		XX	PR	21-MAY-2001; 2001EP-0000166.		
Db	479	TyrGlyAsnPheSerPheThrArgValTyrGlnAlaGlyHisGluValProTyrTyrGln	498		XX	PR	21-MAY-2001; 2001EP-0000168.		
QY	1441	CCCATCGCTCTCTGCAATGTTTAAACGGGACTATCTCGGTGGATATCCGAGAGGC	1500		XX	PR	21-JUN-2001; 2001EP-0000240.		

PR 22-OCT-2001; 2001EP-0000557.  
PR 22-OCT-2001; 2001EP-0000558.  
PR 15-NOV-2001; 2001EP-0004464.  
PR 21-DEC-2001; 2001EP-0005117.

(STAM ) DSM NV.

PI Edens L, Van Dijk MA, Krubasik P, Albersmann K, Stock A, Kimpel E;  
PI Klubbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;  
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;  
PI Stiebler J, Albarg R;

WPI; 2002-723203/78.

P-PSDB; ABZ78232, ABZ78289.

Novel isolated protease polypeptide useful in laboratory, clinical, pharmaceutical, chemical, diagnostic, personal care and industrial applications -

PS Claim 13; Page 382-385; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A polypeptide or polynucleotide of the invention is useful for diagnosing a fungal infection such as aspergillosis, or as a query sequence to perform a search against public databases. A polypeptide of the invention is useful in a selected number of industrial or pharmaceutical processes, in laboratory or clinical processes, in food industry (baking, brewing, cheese manufacturing, meat tenderising), in tanning industry and in the manufacture of biological detergents. A polypeptide may also be useful for improving protein solubility, extraction yields, viscosity or taste, texture, nutritional value, minimising of antigenicity or anti-nutritional factors, colour or functionality as well as processing aspects like filterability of the proteinaceous raw material. The sequences shown in ABR38313-ABR38869 represent the *A. niger* processes of the invention.

Sequence	623 AA:
<p> <del> </del> </p>	<p> <del> </del> </p>

Alignment Scores:

Pred. No.:	2,14e-80	Length:	623
Score:	580.50	Matches:	216
Percent Similarity:	56.23%	Conservative:	91
Local Similarity:	39.56%	Mismatches:	178
Query Match:	33.76%	Indels:	61
DB:	23	Gaps:	19

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x ABR38865 (1-623)

76 ACCGCAACAATGTCACCATCCGGTACAAGGAA-----CCCGGGGCGAG----- 120

32 ThrProGluAspLeuThrValIleHisSerGluIlePheProGlyAlaArgIleSerTyr 51

[27] -----

124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942

[illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

169 GACACCTCTCCCGAGTCC-----186

98T -----

72 HisLeuProProHisThrLeuThrAsnLeuSerIleProGlyIleSerIleSerGlnPro 91

187

187 -----CATACCTTCTCTGTTCTTCGAAGCCAGA-----CATAACCCAGAAACT 231

[illegible]

Y P R C A S M L F N E I E T P Y I P H E P R O S E R A R G H I S H I S A s n A s p T h r I l l

232 GCACCTATCACATTGTGGTTGAATGGTGGCCCTGGAAGCGATTCTTTTGATCGGCTCTCTTC 391

291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 528  
 529  
 530  
 531  
 532  
 533  
 534  
 535  
 536  
 537  
 538  
 539  
 540  
 541  
 542  
 543  
 544  
 545  
 546  
 547  
 548  
 549  
 550  
 551  
 552  
 553  
 554  
 555  
 556  
 557  
 558  
 559  
 560  
 561  
 562  
 563  
 564  
 565  
 566  
 567  
 568  
 569  
 570  
 571  
 572  
 573  
 574  
 575  
 576  
 577  
 578  
 579  
 580  
 581  
 582  
 583  
 584  
 585  
 586  
 587  
 588  
 589  
 590  
 591  
 592  
 593  
 594  
 595  
 596  
 597  
 598  
 599  
 600  
 601  
 602  
 603  
 604  
 605  
 606  
 607  
 608  
 609  
 610  
 611  
 612  
 613  
 614  
 615  
 616  
 617  
 618  
 619  
 620  
 621  
 622  
 623  
 624  
 625  
 626  
 627  
 628  
 629  
 630  
 631  
 632  
 633  
 634  
 635  
 636  
 637  
 638  
 639  
 640  
 641  
 642  
 643  
 644  
 645  
 646  
 647  
 648  
 649  
 650  
 651  
 652  
 653  
 654  
 655  
 656  
 657  
 658  
 659  
 660  
 661  
 662  
 663  
 664  
 665  
 666  
 667  
 668  
 669  
 670  
 671  
 672  
 673  
 674  
 675  
 676  
 677  
 678  
 679  
 680  
 681  
 682  
 683  
 684  
 685  
 686  
 687  
 688  
 689  
 690  
 691  
 692  
 693  
 694  
 695  
 696  
 697  
 698  
 699  
 700  
 701  
 702  
 703  
 704  
 705  
 706  
 707  
 708  
 709  
 710  
 711  
 712  
 713  
 714  
 715  
 716  
 717  
 718  
 719  
 720  
 721  
 722  
 723  
 724  
 725  
 726  
 727  
 728  
 729  
 730  
 731  
 732  
 733  
 734  
 735  
 736  
 737  
 738  
 739  
 740  
 741  
 742  
 743  
 744  
 745  
 746  
 747  
 748  
 749  
 750  
 751  
 752  
 753  
 754  
 755  
 756  
 757  
 758  
 759  
 760  
 761  
 762  
 763  
 764  
 765  
 766  
 767  
 768  
 769  
 770  
 771  
 772  
 773  
 774  
 775  
 776  
 777  
 778  
 779  
 780  
 781  
 782  
 783  
 784  
 785  
 786  
 787  
 788  
 789  
 790  
 791  
 792  
 793  
 794  
 795  
 796  
 797  
 798  
 799  
 800  
 801  
 802

112 SerProLeuThrIleIrpMetAsnGlyGlyproGlyGlySerSerMetIleGlyLeuphe 131

202 CANNONWOOD DRIVE  
202 CANNONWOOD DRIVE  
202 CANNONWOOD DRIVE

292 GAAGAGTTGGGCCCTTGCCATGTCAATTCGACTTTTGATGAC---TACATCAACCCCTCAC 348

[illegible]

152 CINGGDSNGLYPIOCYSINI VALASNTHTASpSerAsnSerThrAlaTyrAsnProTrp 151

Y  
349 TCGTGGACGAGGTCTCCAATTACTATTCCTGTCCAGGCATTGCCA

5'-TGGGAGTCGGCTTTTCA 408

Db	152	SerTrpAsnGluTyrValAspMetLeuTyrIleGluGlnProValGlnThrGlyPheSer	171
QY	409	TATAGTATACCGTGGTGGTCCATTAACCCCTGTAACCTGGGCTGTCGAAATTCGACG	468
Db	172	TyrAspValLeuArgAsnGlyThrLeuAsp	185
QY	469	TTTGCAGGAGTTCAGGCGCGGTACCCAAACCATTTGAT	522
Db	186	Phe	203
QY	523	ACCAATCTTCCGCAGACGCCCGCTGGGAGATCCCTGCAGAGTTCCTTAGTGGGA	576
Db	204	ValAsn	222
QY	577	CTAAGCTTAGCTTGACCTAGGTGGTGCAGTCTAAGGAC	627
Db	223	GluTyrValSerSerValAspGlyAsnGlyGlyGlyAspAspArgValSerIleTrpThr	242
QY	628	GAGAGCTATGGAGGCACTATGGTCTCGATTCATTCATTCATTTTACGAGCAGAAATCAG	687
Db	243	GluSerTyrGlyGlyArgGlyProAlaTyrThrAlaLeuPheGlnGluMetAsnGlu	262
QY	688	AGAAATGCCACGGTAGTGTAAAT	744
Db	263	ArgIleGluSerGlyGluValSerThrGlyLysIleHisLeuAspThrLeuGlyIle	282
QY	745	ATTAAAGGCATCATCGACGGCGATCCAGCGCCCTTACTACCCGGAATTCGCTGTGAC	804
Db	283	IleAsnGlyCysValAspLeuLeuValGlnValProSerPheProGluAlaTyrAsn	302
QY	805	ANTACCTACGGTATCAAGGCTGTCAACGAGACCGCTCTACAACCTACATGAAGTTTGCCAAC	864
Db	303	AsnThrTyrGlyIleGluGlyIleAsnArgThrLeuTyrAspArgAlaMetAspSerTrp	322
QY	865	CAAAATGCCAAATGGTTGCCAGGATTTGATTTCCACCTGTCAACAGACAAACCGCACCGCA	924
Db	323	SerLysProGlyGlyCysArgAspMetIleIleGluCysArgAspAlaGlyGlu	340
QY	925	TTAGCTGAC	981
Db	341	LeuGlyAspProLeuIleIleCysGluGluAlaSerAspTyrCysSerArgGluIleLys	360
QY	982	GGGCATACTACGGCTTTGTGTGGTGGTGTATGATATATCGCATCATCAATGATGAC	1041
Db	361	SerLeuTyrThrAsnThrSerGlyArgGlyTyrTyrAspIleAlaHisPheIhrProAsp	380
QY	1042	CCGACTCCGCCAAGTTATTACACAAATTTCTGCAAGGACTCTGTCACTGGAGGCTATTC	1101
Db	381	AlaAlaLeuValProTyrPheValGlyPheLeuAsnArgProTyrPheValGlnLysAlaLeu	400
QY	1102	GGCTCAACTCACTACACCCAGTCCAAATAGACGCTCTACTACGCTTTCCAGCAACA	1161
Db	401	GlyValProValAsnTyrThrMetSerSerGluAlaValGlyAsnSerPheAlaSerThr	420
QY	1162	GGCAGCTTTGTCTGGGCC	1206
Db	421	GlyAspTyr	438
QY	1207	CTTGCTCTCCCGTGGTGTCTCCCTCACTATGCGGACCGCGATTCACACTCACTGAG	1266
Db	439	LeuAspSerGlyValLysValAlaMetValTyrGlyAspArgAspTyrAlaCysProTyr	458
QY	1267	TTCCGGCGGTCCAGCGGTTTCCCTCGCTGGGAACTACTCCCAACGCCCGCTCCGAGC	1326
Db	459	ArgGlyGlyGluAspValSerLeuLeuValGluTyrGluAspAlaGluLysPheArgAla	478
QY	1327	GCAGGTTACACGCCCTGAAAGTCAACGGCGTCCGAGTAT	1380
Db	479	AlaGlyTyrAlaGluValGlnThrLysSer	497
QY	1381	TATGGTAAATTCCTCTACTCGCGTCTATGAGGCGGACCATGAAGTCCCATCTACCAG	1440

Db 498 TyrGlyAsnPheSerPheThrArgValPheGlnAlaGlyHisGluValProPheThrGln 517  
QY 1441 CCCATCGCTCCCTGCAATGTTTAAACGGGACTATCTTCGTTGGGATATCGCAGAGGC 1500  
Db 518 ProGluThrAlaTyrGluLeuPheAsnArgAlaGlnPheAsnTrpAspIleAlaThrGly 537  
QY 1501 CAGAAGAAGACTGG-----CCCAGCTACAGAGCAATGGAAACGGCTACAGCTACGCAT 1554  
Db 538 GlyIleSerLeuGluGlnAsnGlnSerIyrGlyThrGluGlyProSerSerThrTrpHis 557  
QY 1555 ACACAGTCGTCGTCGCCG 1572  
Db 558 IleLysAsnGluValPro 563  
RESULT 4  
ID ABR38859 standard; Protein; 554 AA.  
XX AC ABR38859;  
XX DT 24-APR-2003 (first entry)  
XX DE A. niger serine carboxypeptidase polypeptide #2.  
XX KW Protease; fungal infection; aspergillosis; food; tanning; detergent;  
KW protein solubility; viscosity; taste; texture; nutritional value;  
KW EC3.4.16.6.  
XX AS Aspergillus niger.  
XX PN WO200268623-A2.  
XX PD 06-SEP-2002.  
XX PF 22-FEB-2002; 2002WO-EF01984.  
XX PR 23-FEB-2001; 2001EP-0200657.  
XX PR 23-FEB-2001; 2001EP-0200658.  
XX PR 23-FEB-2001; 2001EP-0200660.  
XX PR 26-FEB-2001; 2001EP-0200706.  
XX PR 26-FEB-2001; 2001EP-0200707.  
XX PR 26-FEB-2001; 2001EP-0200708.  
XX PR 26-FEB-2001; 2001EP-0200719.  
XX PR 28-MAR-2001; 2001EP-0000075.  
XX PR 28-MAR-2001; 2001EP-0000078.  
XX PR 28-MAR-2001; 2001EP-0000080.  
XX PR 28-MAR-2001; 2001EP-0000087.  
XX PR 28-MAR-2001; 2001EP-0000088.  
XX PR 21-MAY-2001; 2001EP-0000156.  
XX PR 21-MAY-2001; 2001EP-0000159.  
XX PR 21-MAY-2001; 2001EP-0000160.  
XX PR 21-MAY-2001; 2001EP-0000162.  
XX PR 21-MAY-2001; 2001EP-0000165.  
XX PR 21-MAY-2001; 2001EP-0000166.  
XX PR 21-MAY-2001; 2001EP-0000168.  
XX PR 21-JUN-2001; 2001EP-0000240.  
XX PR 21-JUN-2001; 2001EP-0000242.  
XX PR 21-JUN-2001; 2001EP-0000244.  
XX PR 21-JUN-2001; 2001EP-0000246.  
XX PR 12-JUL-2001; 2001EP-0000280.  
XX PR 12-JUL-2001; 2001EP-0000285.  
XX PR 30-JUL-2001; 2001EP-0000323.  
XX PR 30-JUL-2001; 2001EP-0000327.  
XX PR 02-AUG-2001; 2001EP-0000341.  
XX PR 02-AUG-2001; 2001EP-0000342.  
XX PR 02-AUG-2001; 2001EP-0000343.  
XX PR 02-AUG-2001; 2001EP-0000344.  
XX PR 09-AUG-2001; 2001EP-0000357.  
XX PR 16-AUG-2001; 2001EP-0000374.  
XX PR 16-AUG-2001; 2001EP-0000377.  
XX PR 20-SEP-2001; 2001EP-0000478.  
XX PR 20-SEP-2001; 2001EP-0000483.  
XX PR 22-OCT-2001; 2001EP-0000552.

PR 22-OCT-2001; 2001EP-0000553.  
PR 22-OCT-2001; 2001EP-0000554.  
PR 22-OCT-2001; 2001EP-0000556.  
PR 22-OCT-2001; 2001EP-0000557.  
PR 22-OCT-2001; 2001EP-0000558.  
PR 15-NOV-2001; 2001EP-0004464.  
PR 21-DEC-2001; 2001EP-0005117.  
XX (STAM ) DSM NV.  
XX PA Edens L, Van Dijk AA, Krubasik P, Alberman K, Stock A, Kimpel E;  
PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;  
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;  
PI Stiebler J, Albang R;  
XX WPI: 2002-723203/78.  
DR P-PSDB; ABZ78226, ABZ78283.  
DR Novel isolated protease polypeptide useful in laboratory, clinical,  
PT pharmaceutical, chemical, diagnostic, personal care and industrial  
PT applications -  
XX Claim 13; Page 366-369; 394pp; English.  
XX CC The invention relates to a novel isolated protease polypeptide. A  
CC polypeptide or polynucleotide of the invention is useful for diagnosing a  
CC fungal infection such as aspergillosis, or as a query sequence to perform  
CC a search against public databases. A polypeptide of the invention is  
CC useful in a selected number of industrial or pharmaceutical processes, in  
CC laboratory or clinical processes, in food industry (baking, brewing, in  
CC cheese manufacturing, meat tenderising), in tanning industry and in the  
CC manufacture of biological detergents. A polypeptide may also be useful  
CC for improving protein solubility, extraction yields, viscosity or taste,  
CC texture, nutritional value, minimising of antigenicity or  
CC anti-nutritional factors, colour or functionality as well as processing  
CC aspects like filterability of the proteinaceous raw material. The  
CC sequences shown in ABR38859 represent the A. niger proteases of  
CC the invention.  
XX SQ Sequence 554 AA;  
Alignment Scores:  
Pred. No.: 2.84e-77 Length: 554  
Score: 946.00 Matches: 213  
Percent Similarity: 51.72% Conservative: 73  
Best Local Similarity: 38.52% Mismatches: 176  
Query Match: 32.58% Indels: 91  
DB: Gaps: 13  
US-09-712-338-1\_COPY\_55\_1662 (1-1608) x ABR38859 (1-554)  
QY 37 CAGCTACCCCAAGAACCCCGGGTCAAGACTCTTACACCGCA-----ACAAATGTC 90  
Db 17 GlnPheProGluProGluProGluGlyIleThrValLeuLysSerLysLeuHisGluAsnVal 36  
QY 91 ACCATCGGTACAGGACCGCGGCGCAGAGGCGCTCGAGACTACCCCGGGTCTCAA 150  
Db 37 ThrileSerPheLysGluPro-----GlyIleCysGluThrThrProGlyValarg 53  
QY 151 TCCTACTCTGATATGTCGACACCTCTCCCGAGTCCCACTCTCTCTGTTCTCGAA 210  
Db 54 SerTyrSerGlyTyrValHisLeuProAlaSerThrSerPheThrPhePheGlu 73  
QY 211 GCCAGACATAACCCAGAAACCTGCACCTATACATTCGTTGATGGTGGCCCTGGGAGC 270  
Db 74 AlaArgLysAspProSerAsnAlaProLeuAlaIleTrpLeuAsnGlyGlyProGlyGly 93  
QY 271 GATCTTTGATCGGTCTCTTCGAGAGTTGGCCCTTGCATGTCAATTCG---ACTTTT 327  
Db 94 SerSerLeuMetGlyLeuLeuGluGlyProCysSerIleAlaSerAspSerLys 113  
QY 328 GATGACTACATCAACCTCTCACTCGTGTGGAACGAGGTCTCCAAATTTACTATTCTGTCGAG 387

Db 114 ThrThrValLeuAsnProTrpSerTrpAsnAsnGluValAsnLeuLeuPheLeuAspGln 133  
 QY 388 CCATGGAGTGGCTTTTCATATAGTGTACGGTGTGATGGGTCATTAACCCGTGACT 447  
 Db 134 ProThrGlnValGlyPheSerTrpAspValProThrAsnGlyThrLeuThr----- 150  
 QY 448 GGGGTCTCGAAATTCGAGCTTTGCGAGGAGTTTCAGGCGCGGTACCCAAACCAITGATGCC 507  
 Db 150 ----- 150  
 QY 508 ACTCTGATGATACCTACCAATCTTGGCGCAGAGCCCTTGGGAGATCCCTGCAAGGATTC 567  
 Db 151 -----AlaAsnGlyThrAlaPheAlaAlaHisAlaLeuTrpHisPheAlaGlnThrTrp 168  
 QY 568 CTTAGTGGACTACCTAGCTTG-----GACTCTAGGCTGAGTCTAAGGACTTTCAGT 618  
 Db 169 PhePheGluPheProHisTrpLysProAsnAspArgVal-----Ser 183  
 QY 619 CTATGGACGAGAGCTATGGAGGCGACTATGGTCTGTCATCTCTCAATCTTTCATGAG 678  
 Db 184 LeuTrpAlaGluSerTrpGlyGlyHisTrpGlyProGlyLeuPheArgPhePheGlnGln 203  
 QY 679 CAGATGAGAGATTCGCAACGCTAGTGT---AATGGTCTGAG---CTTAATTTCAAC 732  
 Db 204 GlnAsnAspLysLeuAlaGluGlyThrAlaGluAspGlyAlaGlnTrpLeuHisLeuAsp 223  
 QY 733 TCCTCTGGGATATTATTAACGATCATCGAGAGCGGATCCAGCCCTTACTACCTGAA 792  
 Db 224 ThrLeuGlyLeuValAsnGlyLeuMetAspMetValIleGlnGluAlaTrpIleThr 243  
 QY 793 TTCGCTGTCAACAATACCTACGATATCAAGGCTGTCAAGGAGACCGTCTACAACATCAT 852  
 Db 244 TrpPro-----ATGTT-----TyrAsnAsnVal 249  
 QY 853 AAGTTTGCCACCAATGCAATGCTGCCAGGATTTGATTCACCTGCCAACAGACA 912  
 Db 250 ArgLeuAlaPro-----SerSerPhe 256  
 QY 913 AACCGACCGCAATAGCTGACTACGCGCTCTGCGCGAAGCCACCAACATGTGCGAGGAC 972  
 Db 257 AsnSerArgGlyPheArgAspGlnAlaLeuAlaCysGluAlaLeuLysGluArgAsp 276  
 QY 973 -----ATGTT----- 978  
 Db 277 SerGlyLeuProHisSerGlyLysAsnIleSerGluIleCysGlyGlyLeuAlaLeuGlu 296  
 QY 979 -----GAGGGCCA-----TACTACGCTTTGCTGCTGCTGTGTATGATATTCG 1026  
 Db 297 TrpGlyAspGlyProIleThrTrpTrpHisThrPheAsnArgGlyTrpTrpAspIleAla 316  
 QY 1027 CAICCATATGATGACCGGACTCCGCCAAGTATTACACAAATTTCTGCAAGGACTCT 1086  
 Db 317 HisProLysAsnAspProPheProAlaLysHisMetLeuGlyTrpLeuThrGlnGluSer 336  
 QY 1087 GTCATGGAGCTATCGCGCTCAACATCACTACACCCAGTCCAAATAGCATGCTACATAC 1146  
 Db 337 ValLeuAlaAlaLeuGlyValProValAsnPheThrSerSerSerAlaValAlaThr 356  
 QY 1147 GCTTCCAGCAACAGCGGCTTCTGCGCCCACTTCATCGAAGACCTCGAGAGATC 1206  
 Db 357 GlnPheIleLysThrPheAspIleValHisGlyGlyPheLeuAspAlaIleGlyTrpLeu 376  
 QY 1207 CTTGCTCCCGCGTGTCTCCCTCATCTATGCGGACGCGGATTCATCTGCAACTGG 1266  
 Db 377 LeuAspSerGlyValLysValHisMetMetTrpGlyAspArgAspTrpAlaCysAsnTrp 396  
 QY 1267 TTCGCGGTACGGGTTTCCTCGCTCGCAACTACTCCCAAGCGCGGCTCCGGAAGC 1326  
 Db 397 ValGlyGlyGluLysAlaSerLeuAlaValProTrpSerArgIleThrGluPheAlaAsp 416  
 QY 1327 GCAGGTACAGCGCCCTCAAGTCAACGGCTCGAGTATGGGAACTCGCGAGTATGTT 1386  
 Db 417 ThrGlyTrpProLeuLeuThrProAspGlyIleSerGlyMetThrArgGlnLeuGly 436

QY 1387 AATTCTCTCTCTACTCGCTGTATGAGCAGCGCATAGTCCCTACTACTACGCGCCATC 1446  
 Db 437 AsnTrpSerPheThrArgValPheGlnAlaGlyHisGluValProSerTrpGlnProVal 456  
 QY 1447 GCCTCCCTGCAATTTGTTTAAACCGGACTATCTTCGTTGGGTATGCGAGAGGCCAGAAG 1506  
 Db 457 AlaAlaTrpGluIlePheMetArgAlaThrPheAsnLysAspIleProThrGlyLeuLeu 476  
 QY 1507 AAGATCTGCGCCAGCTACAGCAATGGAACGGCTACAGCTACGCTACACAGCTCGTCC 1566  
 Db 477 AlaValAspAspGluPheGlnSerValGlyProLysAspThrTrpHisIleLysAsnIle 496  
 QY 1567 GTGCGCTGCTGCTACCGGTACCGATGTCACAGTGTG 1603  
 Db 497 ProProIle-MetProLysProGlnCysTyrValLeu 508  
 RESULT 5  
 ABR38817  
 ID ABR38817 standard; Protein; 526 AA.  
 XX ABR38817;  
 AC  
 XX  
 DT  
 XX  
 DE  
 XX  
 KW  
 KW  
 KW  
 XX  
 OS  
 XX  
 PN  
 XX  
 PD  
 XX  
 PF  
 XX  
 PR 23-FEB-2001; 2001EP-0200657.  
 PR 23-FEB-2001; 2001EP-0200658.  
 PR 26-FEB-2001; 2001EP-0200706.  
 PR 26-FEB-2001; 2001EP-0200707.  
 PR 26-FEB-2001; 2001EP-0200708.  
 PR 26-FEB-2001; 2001EP-0200719.  
 PR 28-MAR-2001; 2001EP-0000075.  
 PR 28-MAR-2001; 2001EP-0000078.  
 PR 28-MAR-2001; 2001EP-0000080.  
 PR 28-MAR-2001; 2001EP-0000087.  
 PR 28-MAR-2001; 2001EP-0000088.  
 PR 21-MAY-2001; 2001EP-0000156.  
 PR 21-MAY-2001; 2001EP-0000159.  
 PR 21-MAY-2001; 2001EP-0000160.  
 PR 21-MAY-2001; 2001EP-0000162.  
 PR 21-MAY-2001; 2001EP-0000165.  
 PR 21-MAY-2001; 2001EP-0000166.  
 PR 21-MAY-2001; 2001EP-0000168.  
 PR 21-JUN-2001; 2001EP-0000240.  
 PR 21-JUN-2001; 2001EP-0000242.  
 PR 21-JUN-2001; 2001EP-0000244.  
 PR 21-JUN-2001; 2001EP-0000246.  
 PR 12-JUL-2001; 2001EP-0000280.  
 PR 12-JUL-2001; 2001EP-0000285.  
 PR 30-JUL-2001; 2001EP-0000323.  
 PR 30-JUL-2001; 2001EP-0000327.  
 PR 02-AUG-2001; 2001EP-0000341.  
 PR 02-AUG-2001; 2001EP-0000342.  
 PR 02-AUG-2001; 2001EP-0000343.  
 PR 09-AUG-2001; 2001EP-0000344.  
 PR 16-AUG-2001; 2001EP-0000357.  
 PR 16-AUG-2001; 2001EP-0000374.  
 PR 16-AUG-2001; 2001EP-0000377.  
 A. niger serine carboxypeptidase polypeptide #1.  
 Protease; fungal infection; aspergillosis; food; tanning; detergent;  
 protein solubility; viscosity; taste; texture; nutritional value;  
 EC3.4.16.6.  
 Aspergillus niger.  
 W0200268623-A2.  
 06-SEP-2002.  
 22-FEB-2002; 2002WO-EP01984.  
 23-FEB-2001; 2001EP-0200657.  
 23-FEB-2001; 2001EP-0200658.  
 26-FEB-2001; 2001EP-0200706.  
 26-FEB-2001; 2001EP-0200707.  
 26-FEB-2001; 2001EP-0200708.  
 26-FEB-2001; 2001EP-0200719.  
 28-MAR-2001; 2001EP-0000075.  
 28-MAR-2001; 2001EP-0000078.  
 28-MAR-2001; 2001EP-0000080.  
 28-MAR-2001; 2001EP-0000087.  
 28-MAR-2001; 2001EP-0000088.  
 21-MAY-2001; 2001EP-0000156.  
 21-MAY-2001; 2001EP-0000159.  
 21-MAY-2001; 2001EP-0000160.  
 21-MAY-2001; 2001EP-0000162.  
 21-MAY-2001; 2001EP-0000165.  
 21-MAY-2001; 2001EP-0000166.  
 21-MAY-2001; 2001EP-0000168.  
 21-JUN-2001; 2001EP-0000240.  
 21-JUN-2001; 2001EP-0000242.  
 21-JUN-2001; 2001EP-0000244.  
 21-JUN-2001; 2001EP-0000246.  
 12-JUL-2001; 2001EP-0000280.  
 12-JUL-2001; 2001EP-0000285.  
 30-JUL-2001; 2001EP-0000323.  
 30-JUL-2001; 2001EP-0000327.  
 02-AUG-2001; 2001EP-0000341.  
 02-AUG-2001; 2001EP-0000342.  
 02-AUG-2001; 2001EP-0000343.  
 09-AUG-2001; 2001EP-0000344.  
 16-AUG-2001; 2001EP-0000357.  
 16-AUG-2001; 2001EP-0000374.  
 16-AUG-2001; 2001EP-0000377.





QY 766 GGGATCCAGGCCCTTACTACCTGAATTCGCTGTG---AACAAATACCTACCGTATCAAG 822  
|||||  
Db 256 LeuileglnAlaAspTyrGluProMetAlaCysGlyGlyGlyTyrHisProVal 275  
823 GCTGTCAAGAGACCGCTTACAACTACATGAATGTTGCCAACCAATGCAATGGTGC 882  
276 LeuSerSerGluGluCysGluLysMetSerLysAlaAlaGlyArg----- 290  
883 CAGGATTTGATTTCCACCTGCAACAGAAACCGCCACCGCATCTACTGACACGCCCTC 942  
291 -----CysArgArgLeuAsnLysLeuCysTyrAlaSerLysSerSer 304  
943 -----TGCGCCGAGCCACCACTGTGCAGGACAAATGTTGAGGGCCACTACTACGCC 936  
305 LeuProCysLleValAlaThrAlaTyrCysAspSerAlaLeuLeuGluProTyrIleAsn 324  
997 TTTGCTGGTGGTGGTGTATGATATTCGGCAATCCATGATGACCGC----- 1044  
325 ---ThrGlyLeuAsnValTyrAspIleArgGlyProCysGluAspAsnSerThrAspGly 343  
1045 -----ACTCCGCCAAGTTATTACAAACAATTTCTGCCAAGGACTCTGTGCATGGAC 1095  
344 MetCysTyrThrGlyLeuArgTyrValAspGlnTyrMetAsnPheProGluValGlnGlu 363  
1096 GCTATCGCGTCAACATC---AACTACACCCAGTCCCAATATGACGTCTACTACGGTTTC 1152  
364 ThrLeuGlySerAspValHisAsnTyrSerGlyCysAspAsnValPheThrGlyPhe 383  
1153 CAGCAAAACAGGCGACTTGTCTGGCCCACTTCATCGAACACCTCGAGGAGATCCTTGCT 1212  
384 LeuPheThrGlyAspGlySerLysPro---PheGlnGlnTyrIleAlaGluLeuLeuAsn 402  
1213 CTCCCGTGGTCTCTCCCTCATCTATGGCGAGCGGATACATCTCAACTGCTGTCGGC 1272  
403 HisAsnIleProValLeuIleTyrAlaGlyAspLysAspTyrIleCysAsnTrpLeuGly 422  
1273 GGTACGCGCTTCCCTCGCTGGGAACTATCCCAAGCCGCCAGCTTCGGAAGCGAGGG 1332  
423 AsnHisAlaTrpSerAsnGluLeuGluTrpIleAsnLysArgArgTyrGlnArgMet 442  
1333 TACACGCGCC-----CTGAAGTCAACGGCGTGCAGTATGGGAACTCGGAGTATGGT 1386  
443 LeuArgProTrpValSerLysGluThrGlyGlnGluLeuGlyGlnValLysAsnTyrGly 462  
1387 AATTCCTCTACTCCGCTATGAGCGAGCGCATGAAGPCCCATCTACTACCAAGCCCATC 1446  
463 ProPheThrPheLeuArgIleTyrAspAlaGlyHisMetValProTyrAspGlnProGlu 482  
1447 GCCTCCCTGCAATGTTTAAACCGGACTATCTCGGT 1482  
483 AlaSerLeuGluMetValAsnSerTrpIleSerGly 494

RESULT 7

ABR38819

ID ABR38819 standard; Protein: 536 AA.

XX AC ABR38819;

XX DT 24-APR-2003 (first entry)

XX DE A. niger carboxypeptidase Y polypeptide #1.

XX KW Protease; fungal infection; aspergillosis; food; tanning; detergent;  
KW protein solubility; viscosity; taste; texture; nutritional value;  
KW EC3.4.16.5.

XX OS Aspergillus niger.

XX PN WO200268623-A2.

XX PD 06-SEP-2002.

XX XX 22-FEB-2002; 2002WO-EP01984.

XX 23-FEB-2001; 2001EP-0200657.  
PR 23-FEB-2001; 2001EP-0200658.  
PR 23-FEB-2001; 2001EP-0200660.  
PR 26-FEB-2001; 2001EP-0200706.  
PR 26-FEB-2001; 2001EP-0200707.  
PR 26-FEB-2001; 2001EP-0200708.  
PR 26-FEB-2001; 2001EP-0200719.  
PR 28-MAR-2001; 2001EP-0000075.  
PR 28-MAR-2001; 2001EP-0000078.  
PR 28-MAR-2001; 2001EP-0000080.  
PR 28-MAR-2001; 2001EP-0000087.  
PR 28-MAR-2001; 2001EP-0000088.  
PR 21-MAY-2001; 2001EP-0000156.  
PR 21-MAY-2001; 2001EP-0000159.  
PR 21-MAY-2001; 2001EP-0000160.  
PR 21-MAY-2001; 2001EP-0000162.  
PR 21-MAY-2001; 2001EP-0000165.  
PR 21-MAY-2001; 2001EP-0000166.  
PR 21-MAY-2001; 2001EP-0000168.  
PR 21-JUN-2001; 2001EP-0000240.  
PR 21-JUN-2001; 2001EP-0000242.  
PR 21-JUN-2001; 2001EP-0000244.  
PR 21-JUN-2001; 2001EP-0000246.  
PR 12-JUL-2001; 2001EP-0000280.  
PR 12-JUL-2001; 2001EP-0000285.  
PR 30-JUL-2001; 2001EP-0000323.  
PR 30-JUL-2001; 2001EP-0000327.  
PR 02-AUG-2001; 2001EP-0000341.  
PR 02-AUG-2001; 2001EP-0000342.  
PR 02-AUG-2001; 2001EP-0000343.  
PR 02-AUG-2001; 2001EP-0000344.  
PR 09-AUG-2001; 2001EP-0000357.  
PR 16-AUG-2001; 2001EP-0000374.  
PR 16-AUG-2001; 2001EP-0000377.  
PR 20-SEP-2001; 2001EP-0000478.  
PR 20-SEP-2001; 2001EP-0000483.  
PR 22-OCT-2001; 2001EP-0000552.  
PR 22-OCT-2001; 2001EP-0000553.  
PR 22-OCT-2001; 2001EP-0000554.  
PR 22-OCT-2001; 2001EP-0000556.  
PR 22-OCT-2001; 2001EP-0000557.  
PR 22-OCT-2001; 2001EP-0000558.  
PR 15-NOV-2001; 2001EP-0000464.  
PR 21-DEC-2001; 2001EP-00005117.  
(STAM ) DSM NV.

(STAM ) DSM NV.

Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;  
Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;  
Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;  
Stiebeler J, Albang R;

WPI: 2002-723203/78.

P-PSDB; ABZ78186, ABZ78243.

Novel isolated protease polypeptide useful in laboratory, clinical,  
pharmaceutical, chemical, diagnostic, personal care and industrial  
applications

Claim 13; Page 279-281; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A  
polypeptide or polynucleotide of the invention is useful for diagnosing a  
fungal infection such as aspergillosis, or as a query sequence to perform  
a search against public databases. A polypeptide of the invention is  
useful in a selected number of industrial or pharmaceutical processes, in  
laboratory or clinical processes, in food industry (baking, brewing,  
cheese manufacturing, meat tenderising), in tanning industry and in the  
manufacture of biological detergents. A polypeptide may also be useful  
for improving protein solubility, extraction yields, viscosity or taste,  
texture, nutritional value, minimising of antigenicity or  
anti-nutritional factors, colour or functionality as well as processing



Db 342 SerAla-----TrpValCysValProAlaSerIleTyrCysAsnAsnAla 356  
QY 976 GTTCAGGGGCATACTACGCGTTTCGTGGTGGTGATGATATTCGGCAATCCATAT 1035  
Db 357 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArgGlyLysCys 375  
QY 1036 GATGACCCG-----ACTCCGCCAAGTTATTACAACAATAATTCITGGCAAAG 1080  
Db 376 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 395  
QY 1081 GACICTGTCAFGACGCTATCGCGTCAACATCAAC---TACACCAGTCCCATATGAC 1137  
Db 396 ProGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp 415  
QY 1138 GTCTACTACGCTTTCCAGCAAAACGCGACTTTGTCTGGCCC-----AACTTCATCGAA 1191  
Db 416 IleAsnArgAsnPheLeuPheHisGlyAspTyrMetLysProTyrHisArgLeuValPro 435  
QY 1192 GAECTCGAGAGATCCCTGTCTCCCGTCGCGTCTCCCTCATCTAT---GGCGACGCC 1248  
Db 436 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla 450  
QY 1249 GATTACATCGCAACTCGTTCGGCGTCAAGCGCGTTCCCTCGCTCGGAAGTAATGAG 1308  
Db 451 AspPheIleCysasnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpProGly 470  
QY 1309 GCAGCGCCAGTTCGGAAGCGAGGTTACAGCCCTGAAAAGTC-----AAC 1353  
Db 471 GlnAlaGluTyrAlaSerAlaGluLeuGluAspLeuValIleValAspAsnGluHisThr 490  
QY 1354 GCGCTCGAGTATGGGAAACTCCGCGTAGTAGTAATTCCTCCCTCACTCGCGTCTAAGAG 1413  
Db 491 GlyLysLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly 510  
QY 1414 GCAGGCGCATGAAGTCCCATACACAGCCATCGCGCTCCCTGCCAATGTTCACCGGACT 1473  
Db 511 GlyGlyHisMetValProMetAspGlnProGluSerLeuGluPhePheAsnArgTrp 530  
QY 1474 ATCTCCGTT---TGG 1485  
Db 531 Leuglyglyglutrp 535

RESULT 8  
AAR96737  
ID AAR96737 standard; Protein; 556 AA.  
XX AAR96737;  
XX AC AC AC  
XX XX XX  
DT 14-AUG-1996 (first entry)  
XX XX XX  
DE A. niger Bo-1 carboxypeptidase Y.  
XX Carboxypeptidase Y; CPY; ascomycete; deuteromycete; host cell;  
KW protease deficiency.  
KW Aspergillus niger strain Bo-1.  
OS OS  
XX WC9609397-Al.  
PN XX  
XX XX  
PD 28-MAR-1996.  
XX XX  
PF 19-SEP-1995; 95WO-US11945.  
XX XX  
PP 20-SEP-1994; 94US-0309341.  
XX XX  
XX (NOVO ) NOVO NORDISK BIOTECH INC.  
XX Thompson SA, Yaver DS;  
XX WPI; 1996-188458/19.  
DR N-PSDB; AAT28283.  
XX XX

PI Nucleic acid construct encoding a filamentous ascomycete or  
 PT deuteromycete carboxypeptidase Y - useful to produce host cells  
 PT modified to produce reduced amounts of carboxypeptidase  
 XX  
 XX  
 XX Claim 4; Fig 1A-E; 46pp; English.  
 PS  
 CC Carboxypeptidase Y (AAR96737), a vacuolar protease, is the product of  
 CC the CPY gene (AAR28283) of *Aspergillus niger* strain Bo-1. Creation  
 CC of CPY-deficient *Aspergillus* strains, e.g. by cloning a selectable  
 CC marker into the CPY gene, provides suitable host strains for prodn.  
 CC of heterologous protein.  
 XX  
 XX Sequence 556 AA;  
 SQ

Alignment Scores:  
 Pred. No.: 1,95e-48 Length: 556  
 Score: 630.00 Matches: 154  
 Percent Similarity: 48.45% Conservative: 81  
 Best Local Similarity: 31.75% Mismatches: 174  
 Query Match: 21.69% Indels: 76  
 DB: 17 Gaps: 17

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x AAR96737 (1-556)

QY 79 GCAAAACATGTCACCATCGGTACAGGAACCCGGGCGAGGGGCTGCGAGACTAC 138  
 DB 131 AlaTyrAspLeuArgValLysThrAspProGlySerLeuGlyLe-----Asp 147  
 QY 139 CGGGGTGTCAAATCCTACTCTGGATATGTCGACACCTCTCCC---GAGTCCCATCTTC 195  
 DB 148 ProGlyValLysGlnThrGlyTyrLeuAspAsnGluAsnAspLysHisLeuPhe 167  
 QY 196 TTTCTGTTCTTGGAGCCAGCATACCAACCCAGAACTGCACCTATCATTTGGTGAAT 255  
 DB 168 TyrTrpPheGluSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 187  
 QY 256 GTGGCCCTGGAGCGATTTTGTATGCTCTTCGAAAGATGGGCGCTTGCAGTIC 315  
 DB 188 GlyGlyProGlyCysSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 207  
 QY 316 AATCGACTTTGATGACTACATCAACCTCTACTCTGTAACGAGCTCTCCATTTACTA 375  
 DB 208 AsnLysLysIleGlnProValTyrAsnAspTyrAlaTrpAsnSerAsnAlaSerValIle 227  
 QY 376 TTTCTCTCCAGCATTTGGAGCTGCTTTTCATATAGTATGATACGGTGTGATGGTCCATT 435  
 DB 228 PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla-----243  
 QY 436 AACCTGTAACTGGGGTGGTCGAAATTCGAGCTTTTCAGGAGTTTCAGGGCCGGTACCCA 495  
 DB 243 -----243  
 QY 496 ACCATTGATGCCACTGTGATCGATACCTACCAATCTTGGCCGAGAGCGGCTTGGGATC 555  
 DB 244 -----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu 257  
 QY 556 CTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGAGTCTAGGCTGAGTCAAGGAGCTTC 615  
 DB 258 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaLysGlnAspPhe 273  
 QY 616 AGTCTATGGAGGAGACTATGAGGGGCACTATGCTCCTGCATTCATCAATCAATTTTAC 675  
 DB 274 HisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeu 293  
 QY 676 GACCAAGATCAGACAATTGCCAAGGTAGTGTATANTGGTCTTCAGCTTAATTTCACTCT 735  
 DB 294 SerHisLysLysArg-----AsnIleAsnLeuGlnSer 304  
 QY 736 CTGGGAATTAATACGGCATCATCGAGAGCGCATCCAGCCGCTTACTACCTCGAATTC 795  
 DB 305 ValLeuIleGlyAsnGlyLeuThrAspGlyTyrThrGlnTyrGluTyrTyrArgProMet 324  
 QY 796 GCTGTGAACAATACCTACGGTATACAGGCTGTCAAGAGACCGTCTCAACTACATCAAG 855

DB 325 AlaCysGlyAspGlyGlyTyrProAlaValLeuAspGluSerSerCysGlnSerMetAsp 344  
 QY 856 TTTGCCAACCAATGCAATGTTCCAGGATTTGATTTCACCTCCAAACAGACAAAC 915  
 DB 345 -----AsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerGlu 361  
 QY 916 CGCACCGCATTAGCTGACTACGCCCTCTGCGCCGGAAGCCACCAACATGTGCAGGGACAAT 975  
 DB 362 SerAla-----TrpValCysValProAlaSerIleTyrCysAsnAla 376  
 QY 976 GTTGAGGGCCATACACTACGCCCTTGTGCTGGTGTGTATGATCAATTCGGCATCATAT 1035  
 DB 377 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArgGlyLysCys 395  
 QY 1036 GATGACCCG-----ACTCCGCCAAAGTTATTACAAACAAATTTCTGCGAAAG 1080  
 DB 396 GluAspSerSerAsnLeuLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 415  
 QY 1081 GACTCTGTATGACGACTATCGCGCTCAACATCAAC---TACACCCAGTCCCAATATGAC 1137  
 DB 416 ProGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp 435  
 QY 1138 GTCTACTACGCTTCCAGCAACAGCGGACTTTGTCTGCCCC-----AACTTCATCGAA 1191  
 DB 436 IleAsnArgAsnPheLeuPheHisGlyAspTrpMetLysProTyrHisArgLeuValPro 455  
 QY 1192 GACCTCGAGGAGATCTTGTCTCTCCCGTGGTGTCTCCCTCATCAT---GGCGAGGCC 1248  
 DB 456 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla 470  
 QY 1249 GATTACATCTCAACTGGTTCGGCGGTGAGCGGCTTCCCTCGCTCGGAACCTACTCCCAA 1308  
 DB 471 AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTyrProGly 490  
 QY 1309 GCGCCCGCATTCGAGCCGAGGTACACGCCCTCGAAAGTC-----AAC 1353  
 DB 491 GlnAlaGluTyrAlaSerAlaGluLeuGluAspLeuValIleValAspAsnGluHisThr 510  
 QY 1354 GCGCTCGAGTATGGGAACTCGGAGTATGTAATTCTCTTCACTCGCTGTATGAG 1413  
 DB 511 GlyLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly 530  
 QY 1414 GCAGGCATGAATGCCATCACTACAGCCCATCGCTCCCTCGCAATTTGTTAACCGGACT 1473  
 DB 531 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPhePheAsnArgTrp 550  
 QY 1474 ATCTTCGGT---TGG 1485  
 DB 551 LeuGlyGlyGluTrp 555

RESULT 9  
 AAR96738  
 ID AAR96738 standard; Protein; 557 AA.  
 XX  
 AC AAR96738;  
 XX  
 XX 14-AUG-1996 (first entry)  
 DT  
 XX  
 XX A. niger SFAG 2 carboxypeptidase Y.  
 DE  
 XX Carboxypeptidase Y; CPY; ascomycete; deuteromycete; host cell;  
 KW protease deficiency.  
 KW  
 XX Aspergillus niger strain SFAG 2.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..138  
 FT /label= Pre-pro\_peptide  
 FT Protein 139..557  
 FI /label= Mat\_protein  
 XX  
 PN W09609397-Al.

XX PD 28-MAR-1996.  
 XX PF 19-SEP-1995; 95WO-US11945.  
 XX PR 20-SEP-1994; 94US-0309341.  
 XX FA (NOVO ) NOVO NORDISK BIOTECH INC.  
 XX PI Thompson SA, Yaver DS;  
 XX DR WPI; 1996-188458/19.  
 XX N-PSDB; AAT28284.  
 XX  
 Nucleic acid construct encoding a filamentous ascomycete or  
 PT deuteromycete carboxypeptidase Y - useful to produce host cells  
 PT modified to produce reduced amounts of carboxypeptidase  
 XX  
 PS Claim 5; Page 23-25; 46pp; English.  
 XX  
 CC Carboxypeptidase Y (AAR96738), a vacuolar protease, is the product of  
 CC the CPY gene (AAT28284) of *Aspergillus niger* strain SFAG 2. Creation  
 CC of CPY-deficient *Aspergillus* strains, e.g. by cloning a selectable  
 CC marker into the CPY gene, provides suitable host strains for prodn.  
 CC of heterologous protein.  
 XX  
 SQ Sequence 557 AA;  
 Alignment Scores:  
 Pred. No.: 1.6e-47 Length: 557  
 Score: 620.00 Matches: 154  
 Percent Similarity: 48.25% Conservative: 80  
 Best Local Similarity: 31.75% Mismatches: 175  
 Query Match: 21.35% Indels: 76  
 DB: 17 Gaps: 18  
 US-09-712-338-1\_COPY\_55\_1662 (1-1608) x AAR96738 (1-557)  
 QY 79 GCAACAATGTCACCATCGGTACAGGAACCGGGCAGAGCGCTGCGAGACTACC 138  
 DB 132 AlaTyrAspLeuArgValLysLysThrAspProSerSerLeuGlyIle-----Asp 148  
 QY 139 CCGGGTGTCAAACTCTACTCTGGATATGTCGACACCTCTCC-----GAGTCCCATACCTTC 195  
 DB 149 ProGlyValLysGlnTyrThrGlyTyrLeuAspAspAsnGluAsnAspLysHisLeuPhe 168  
 QY 196 TTCTGGTTCCTCGAGCCACATACCCAGAACTGACATATCATGTTGGTTGAAT 255  
 DB 169 TyrIrrPheGluSerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsn 188  
 QY 256 GGIGGCCCTGGAAGCGATTCTTTCATCGGTCTCTCGAAGAGTTGGCCCTTGCCATGTC 315  
 DB 189 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 208  
 QY 316 AATTCGACTTTTGTATGACTACATCAACCTCTACTCTGGAACGAGGTCTCCAAATTACTA 375  
 DB 209 AsnLysLysIleGlnProValTyrAsnAspTyrAlaIrrAsnSerAsnAlaSerValle 228  
 QY 376 TTCTGTCCCGAGCATTTGGAGTCGGCTTTTCATATAGTATACAGTTGATGGTTCATT 435  
 DB 229 PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla----- 244  
 QY 436 AACCCCTGTAACCTGGGGTCTGCGAAATTCGAGCTTTCAGAGTTTCAGGCGCGGTACCCA 495  
 DB 244 ----- 244  
 QY 496 ACCATTGATGCCACTCTGATCGATACATPACCANTTTGCCGAGAGCGCGTCTGGAGATC 555  
 DB 245 -----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu 258  
 QY 556 CTCGAGGATTCCTTAGTGAGTACCTAGCTAGCTGGACTCTGAGGTGCGAGTCTAAGGACTTC 615  
 DB 259 LeuThrLeuPheLysGlnPhePro-----GluTyrAlaLysGlnAspPhe 274

QY 616 AGTCTATGACGGAGAGACTATGGAGGCACTATGTGTCCTGCAATTTCTCAATCATTTTAC 675  
 DB 275 HisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeu 294  
 QY 676 GAGCAGAATGAGAGAAATGCCAACGCTAGTGTATGTTGTTGTTGTTGTTGTTGTTGTT 735  
 DB 295 SerHisLysLysArg-----AsnIleAsnLeuGlnSer 305  
 QY 736 CTGGGAATATTATACGGCATCATCAGCAGGCGATCCAGGCCCTTACTACCTCAATTC 795  
 DB 306 ValLeuIleGlyAsnGlyLeuThrAspGlyLeuThrGluTyrGluTyrArgProMet 325  
 QY 796 GCTGTGAACATACCTACGCTGATCAAGGCTGTCAAGCAGACGCTGTCAACTACATGAAG 855  
 DB 326 AlaCysGlyAspGly---GlyTyrProAlaValLeuAspGlu---GlySerCysGlnAla 343  
 QY 856 TTTCGCAACCAATGCCAAATGGTGGCCAGGATTTGATTTCCACCTGCAACAGACAAC 915  
 DB 344 MetAspAsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerSerGlu 362  
 QY 916 CGCACCGCATTAGCTGACTACGCCCTCTGCGCCGAGCACCACATGTGCGAGGCAAT 975  
 DB 363 SerAla-----TrpValCysValProAlaSerIleTyrCysAsnAla 377  
 QY 976 GTTCAGGGGCCATACTACGCCCTTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1035  
 DB 378 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArgGlyLysCys 396  
 QY 1036 GATGACCGG-----ACTCGCCCAAGTTATTACACAAATTTCTGCGCAAG 1080  
 DB 397 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 416  
 QY 1081 GACTCTGTATGAGCGCTATCGCGTCAACATCAAC---TACACCCAGTCCCAATTAATGAC 1137  
 DB 417 ThrGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp 436  
 QY 1138 GTCTACTACGCTTTCCAGCAACAGGCGGACTTTGTCTGGCC-----AATTCAICGAA 1191  
 DB 437 IleAsnArgAsnPheLeuPheHisGlyAspTrpMetLysProTyrHisArgLeuValPro 456  
 QY 1192 GACCTCGAGGAGACTCTTGTCTCCCGTGTCTCCCTCATCTAT---GGCGACGCC 1248  
 DB 457 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla 471  
 QY 1249 GATTACATCTCAACTGTTGCGCGGTCAAGCCGCTTTCCTCTGCGGAGTACTCCCAA 1308  
 DB 472 AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpProGly 491  
 QY 1309 GCCGCCAGTTCGCAAGCGGAGGTACAGCCCTCGAAGTC-----AAC 1353  
 DB 492 GlnAlaGluTyrAlaSerAlaLysLeuGluAspLeuValValGluAsnGluHisLys 511  
 QY 1354 GCGCTCGAGTATGGGAACTCGCGAGTATGTAATTTCTCTTCCCTGCGGTCTATGAG 1413  
 DB 512 GlyLysLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly 531  
 QY 1414 CGAGGCAATGAATCCCATACTACAGCCCATCCCTCCCTGCAATTTGTTTAAGCGGACT 1473  
 DB 532 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPheAsnArgTrp 551  
 QY 1474 ATCTTCGGT---TGG 1485  
 DB 552 LeuGlyGlyGluTrp 556  
 RESULT 10  
 AAR48059  
 ID AAR48059 standard; Protein: 491 AA.  
 XX  
 AC AAR48059;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 20-JUL-1994 (first entry)

XX Sequence of protease C encoded by gene K1.PRC1.  
DE  
XX Protease; yeast; proteolysis.  
KW  
XX Kluyveromyces lactis.  
OS  
XX WO9400579-A1.  
PN  
XX 06-JAN-1994.  
PD  
XX 23-JUN-1993; 93WO-FR00623.  
PF  
XX 25-JUN-1992; 92FR-0007785.  
PR  
XX (RHON ) RHONE POULENC RORER SA.  
PA  
XX Fleer R, Fournier A, Yeh P;  
PI  
XX WPI; 1994-026215/03.  
DR  
XX N-PSDB; AAQ55347.  
DR  
XX New Kluyveromyces yeast with modified protease gene - esp. used  
PT for high yield prodn. of recombinant protein, also DNA encoding  
PT yeast protease and derived peptide(s)  
PT  
XX Disclosure; Page 28-31; 49pp; English.  
PS  
XX The protease gene is to be modified in order to render it (partially)  
CC incapable of producing the natural protein; or result in a non-  
CC functional protease or in a protease with modified proteolytic  
CC activity. The modifications can be introduced in vitro or in situ by  
CC standard genetic engineering techniques or by exposure to mutagenic  
CC agents.  
CC (Updated on 25-MAR-2003 to correct PN field.)  
CC  
XX Sequence 491 AA;  
SQ

Alignment Scores:  
Pred. No.: 7.52e-43 Length: 491  
Score: 568.50 Matches: 151  
Percent Similarity: 44.65% Conservative: 66  
Best Local Similarity: 31.07% Mismatches: 184  
Query Match: 19.58% Indels: 85  
DB: 15 Gaps: 14

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x AAR48059 (1-491)

Qy 79 GCAACAATGTCACCATCCGGTACAGAGAACCCGGGCGAGCGGTGTCGAGACATACC 138  
Db 62 AlatyrsSerLeuAArgLleLysProLeuAspProLysSerLeuGlyVal----- 77  
Qy 139 CCGGGTGCRAATCCTACTCTGGATATGTCACACCTCTCCGAGTCCCATACCTCTCTC 198  
Db 78 AspThrValLysGlnTrpSerGlyTyrLeuAspTyrGlnAspSerLysHisPhePheTyr 97  
Qy 199 TGGTTCCTCGAAGCCAGACATACCCAGAACTGCACCTATCACAATTGCTTGAATGGT 258  
Db 98 TrpPheGluSerArgAsnAspProGluAsnAspProValLleLeuTrpLeuAsnGly 117  
Qy 259 GGCCCTCGAAGCGATTCCTTTCATCGGCTCTCTTCGAGAGAGTGGGCCCTTGCATGTCAAT 318  
Db 118 GlyProGlyCysSerSerPheValGlyLeuPheGluLeuGlyProSerSerLleGly 137  
Qy 319 TCGACTTTTGATGACTACATCAACCTCCTCGTGGACGAGGTCTCCAAATTTACTATTC 378  
Db 138 AlaAspLeuLysProLleTyrAsnProTyrSerTrpAsnSerAsnAlaSerValLlePhe 157  
Qy 379 CTGTCGCCAGCATTTGGAGTGGCTTTTCATATAGTATGATCGGTGTGATGCTCCATTAAC 438  
Db 158 LeuAspGlnProValGlyValGlyPheSerTyrGlyAsp----- 170  
Qy 439 CCTTACTGGGGTCGCGAAATTCGAGCTTTCGAGGAGTTCAGGGCCGGTACCCAAACC 498

Db 170 ----- 170  
Qy 499 ATTGATGCCACTCTGATCGATACCTACCAATCTTCCGCGAGAGCGGCTTGGAGATCCTG 558  
Db 171 -----SerLysValSerThrThrAspAlaLysAspValTyrLlePheLeu 187  
Qy 559 CAGGATTTCCTTAGTGGACTACCTAGCTAGCTTAGGGTCAGCTTAAGCACTTCAGT 618  
Db 188 AspLeuPhePheGluArgPheProHisLeu-----ArgAsnAsnAspPheHis 203  
Qy 619 CTATGACGGAGAGCTATGAGGCGACTATGGTCTCGATCTTCAATCATCTTTTACGAG 678  
Db 204 IleSerGlyGluSerTyrAlaGlyHisTyrLeuProLysLleAlaHis----- 219  
Qy 679 CAGATGAGAGAAATTGCCAACGGTAGTGTAAATGGTGTTCAGCTTAATTTCACTCTCTG 738  
Db 220 -----GluIleAlaValValHisAlaGluAspSerSerPheAsnLeuSerVal 236  
Qy 739 GGAATTATTAACGGCATCATCGACGAGCGCATCCAGGCCCTTACTACCTCAATTCGCT 798  
Db 237 LeuIleGlyAsnGlyPheThrAspProLeuThrGlnTyrGlnTyrGluProMetAla 256  
Qy 799 GTGAACAATACCTACGGTATCAAGCTGTCAACGAGCCGCTCTACACTACATCAAGTTT 858  
Db 257 CysGlyGluGly---GlyTyrProAlaValLeuGlu----- 267  
Qy 859 GCCAACCAAAATGCCAAATGGTTCAGGAT----- 888  
Db 268 -----ProGluAspCysLeuAspMetAsnArgAsnLeuProLeuCysLeuSer 283  
Qy 889 TTGATTTCCACTGCAACAGACAAACCGCACCGCATTTAGCTAGCTACGCCCTCTGCCCC 948  
Db 284 LeuValAspArgCysTyrLysSerHis-----SerValPheSerCysVal 298  
Qy 949 GAAGCCACCAACATGTGCAGGACAAATGTTGAGGGGCCATCTACCGCTTTCGTGCTGT 1008  
Db 299 LeuAlaAspArgTyrCysGluGlnGlnIleThrGly---ValTyrGluLysSerGlyArg 317  
Qy 1009 GGTGTGTATGATATTCGG-----CATCCATATGATGACCCG-----ACT 1047  
Db 318 AsnProTyrAspIleArgSerLysCysGluAlaGluAspAspSerGlyValaCysTyrGln 337  
Qy 1048 CCGCCAAAGTTATTACACAAATTTCTGCCAAGGACTCTGTCAATGAGCGCTATCGCGCTC 1107  
Db 338 GluGluIleTyrIleSerAspTyrLeuAsnGlnGluValGlnArgAlaLeuGlyThr 357  
Qy 1108 AACATCACTACACCCAGTCC---AATAATGACGTCTACTAGCTTTCCAGCAACAGGC 1164  
Db 358 AspValSerSerPheGlnGlyCysSerSerAspValGlyIleGlyPheAlaPheThrGly 377  
Qy 1165 GACTTTGTCTGGCCCAACTTCATCGAGACCTCGAGAGATCCTTGTCTCTCCCGTGGGT 1224  
Db 378 AspGlyProSerPro---PheHisGlnTyrValAlaGluLeuLeuAspGlnAspIleAsn 396  
Qy 1225 GTCTCCCTCATCTATGCGACGCCCATATCATCTGCACTGTCACACTGTTCCGGCGGTCAGCCGTT 1284  
Db 397 ValLeuIleTyrAlaGlyAspLysAspTyrIleCysAsnTrpLeuGlyAsnLeuAlaTrp 416  
Qy 1285 TCCCTCGCTCGGAACACTACTCCCAAGCCGCCAGTTCGGAAGCGCAGGTACAGCCCGCTG 1344  
Db 417 ThrGluLysLeuGlnTrpArgTyrAsnGluGlnTyrLysLysGlnValLeuArgThrTrp 436  
Qy 1345 AAAGTCAACGGCGTCGAG-----TATGGGAACTCCGAGATGTAATTTTCCTTC 1398  
Db 437 LysSerGluGluThrAspGluThrIleGlyGluThrLysSerTyrGlyProLeuThrTrp 456  
Qy 1399 ACTCCGCTCTNAGGCGAGGCATCAAGTCCCATCTACCCCATCGCTCCCTCCGCAA 1458  
Db 457 LeuArgIleTyrAspAlaGlyHisMetValProHisAspGlnProGluAsnSerLeuGln 476  
Qy 1459 TGTGTTTAACCGGACTATC 1476  
Db 1459 TGTGTTTAACCGGACTATC 1476



```
Db      398 AsnLeuAspTyrValIysGluAlaValGlyAlaSerAsnIleAspIlePheThrSerCys 417
QY      1129 ATAATGACGTCTACTACGCTTCCAGAAACAGCGAGCTTGTCTGGCCCAACTTCATC 1188
Db      418 AspAspThrValPheArgAsnPhelIleLeuAspGlyAspGluMetIysPro---PheGln 436
QY      1189 GAAGACCTCGAGGAGATCTTGTCTCCCGTGGTGTCTCCCTCATCTATGCGAGGCC 1248
Db      437 GlnTyrValAlaGluLeuLeuAspAsnValProValLeuIleTyrAlaGlyAspIys 456
QY      1249 GATTACATCGCAACGTGTGGCGGTGAGCGGTTCCTCGCTCGGCAACTACTCCCAA 1308
Db      457 AspTyrIleCysAsnTrpLeuGlyAsnLeuAlaTrpValAsnGluLeuGluTyrSerAsp 476
QY      1309 GCGGCCAGTTCGGAAGCGCAGGTACAGCCCTGAAAGTCAACGCGCTCGAG--TAT 1365
Db      477 SerGluHisPheAlaProIysProLeuGlnLeuTrpIysGlnAspGlyIysAlaAla 496
QY      1366 GGGGAACACTCGGAGTATGTAATTTCTCTTCACCTCGGTCTATGAGCGAGCCATGAA 1425
Db      497 GlyGluValIysAsnHisIysHisPheThrPheLeuArgIleTyrAspAlaGlyHisMet 516
QY      1426 GTCCCATACTACAGCCCTCGCTCCCTGCAATTTGTTAAACCGGACTATCTTCGTT 1482
Db      517 ValProPheAspGlnProGluAsnAlaLeuSerMetValAsnThrTrpValGlnGly 535

RESULT 12
ABR38845
ID   ABR38845 standard; Protein; 481 AA.
XX
AC   ABR38845;
XX
DT   24-APR-2003 (first entry)
XX
DE   A. niger carboxypeptidase Y polypeptide #3.
XX
KW   Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW   protein solubility; viscosity; taste; texture; nutritional value;
KW   EC3.14.16.5.
XX
OS   Aspergillus niger.
XX
PN   WO200268623-A2.
PD
PF   06-SEP-2002.
PF   22-FEB-2002; 2002WO-EP01984.
XX
PR   23-FEB-2001; 2001EP-0200657.
PR   23-FEB-2001; 2001EP-0200658.
PR   23-FEB-2001; 2001EP-0200660.
PR   26-FEB-2001; 2001EP-0200706.
PR   26-FEB-2001; 2001EP-0200707.
PR   26-FEB-2001; 2001EP-0200708.
PR   26-FEB-2001; 2001EP-0200719.
PR   28-MAR-2001; 2001EP-0000075.
PR   28-MAR-2001; 2001EP-0000078.
PR   28-MAR-2001; 2001EP-0000080.
PR   28-MAR-2001; 2001EP-0000087.
PR   28-MAR-2001; 2001EP-0000088.
PR   28-MAR-2001; 2001EP-0000156.
PR   21-MAY-2001; 2001EP-0000159.
PR   21-MAY-2001; 2001EP-0000160.
PR   21-MAY-2001; 2001EP-0000162.
PR   21-MAY-2001; 2001EP-0000165.
PR   21-MAY-2001; 2001EP-0000166.
PR   21-MAY-2001; 2001EP-0000168.
PR   21-JUN-2001; 2001EP-0000240.
PR   21-JUN-2001; 2001EP-0000242.
PR   21-JUN-2001; 2001EP-0000244.
PR   21-JUN-2001; 2001EP-0000246.
PR   12-JUL-2001; 2001EP-0000280.
PR   12-JUL-2001; 2001EP-0000285.

PR      30-JUL-2001; 2001EP-0000323.
PR      30-JUL-2001; 2001EP-0000327.
PR      02-AUG-2001; 2001EP-0000341.
PR      02-AUG-2001; 2001EP-0000342.
PR      02-AUG-2001; 2001EP-0000343.
PR      02-AUG-2001; 2001EP-0000344.
PR      09-AUG-2001; 2001EP-0000357.
PR      16-AUG-2001; 2001EP-0000374.
PR      16-AUG-2001; 2001EP-0000377.
PR      20-SEP-2001; 2001EP-0000478.
PR      22-OCT-2001; 2001EP-0000552.
PR      22-OCT-2001; 2001EP-0000553.
PR      22-OCT-2001; 2001EP-0000554.
PR      22-OCT-2001; 2001EP-0000556.
PR      22-OCT-2001; 2001EP-0000557.
PR      22-OCT-2001; 2001EP-0000558.
PR      15-NOV-2001; 2001EP-0004454.
PR      21-DEC-2001; 2001EP-0005117.
XX
XX      (STAM ) DSM NV.
XX
PI      Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
PI      Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
PI      Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
PI      Stiebler J, Albarg R;
XX
XX      WPI; 2002-723203/78.
DR      P-PSDB; ABZ78212, ABZ78269.
XX
XX      NDB isolated protease polypeptide useful in laboratory, clinical,
PT      pharmaceutical, chemical, diagnostic, personal care and industrial
PT      applications
XX
XX      Claim 13; Page 335-337; 394pp; English.
XX
XX      The invention relates to a novel isolated protease polypeptide. A
CC      polypeptide or polynucleotide of the invention is useful for diagnosing a
CC      fungal infection such as aspergillosis, or as a query sequence to perform
CC      a search against public databases. A polypeptide of the invention is
CC      useful in a selected number of industrial or pharmaceutical processes, in
CC      laboratory or clinical processes, in food industry (baking, brewing, in
CC      cheese manufacturing, meat tenderising), in tanning industry and in the
CC      manufacture of biological detergents. A polypeptide may also be useful
CC      for improving protein solubility, extraction yields, viscosity or taste,
CC      texture, nutritional value, minimising of antigenicity or
CC      anti-nutritional factors, colour or functionality as well as processing
CC      aspects like filterability of the proteinaceous raw material. The
CC      sequences shown in ABR38313-ABR38869 represent the A. niger proteases of
CC      the invention.
XX
XX      Sequence 481 AA;
SQ
XX
XX      Alignment Scores:
XX      Pred. NO.: 9.35e-29 Length: 481
XX      Score: 414.00 Matches: 140
XX      Percent Similarity: 41.73% Conservative: 77
XX      Best Local Similarity: 26.92% Mismatches: 207
XX      Query Match: 14.26% Indels: 96
XX      DB: 23 Gaps: 20
XX
XX      US-09-712-338-1_COPY_55_1662 (1-1608) x ABR38845 (1-481)
QY      43 CCCAAGAACCCACCGGGGTCAAGACTCTT-----ACAACGCAACAATGTC 90
Db      23 ProGluGluProSerAspPheArgThrPheHisSerProTyrSerProHisHisSerIle 42
QY      91 ACCATCGGTACACAGGACCCGCGGCGAGCGGCTCTGCGAGACTACCCCGGTGTCAAA 150
Db      43 ArgIleArgGlnGlnAsn-----GluSerIleCysAlaAlaHisSerAla----- 57
QY      151 TCCTACTCTGATATGTGACACACCTCTCCCGAGTCCCATACCTCTTCTGGTTCGAA 210
```



PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160983.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26 OCT-1999; 99US-0161361.



[illegible]

PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 6,43e-26  
 Score: 383.00  
 Percent Similarity: 39.42%  
 Best Local Similarity: 26.21%  
 Query Match: 13.19%  
 DB: 21  
 Gaps: 18

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x AAG30064 (1-502)

QY 43 CCCAAGACCCGACCGGGTCAAGACTCTTACAAACCGCAAAATGTCAACATCGGTAC 102  
 DB 52 ProSerProLysArgGlyValSerSerSerGlyAspThrSerHisPheAsnValIleGln 71  
 QY 103 AAGGAA-----CCCGGGCAGAGGCGTCTCGAGACTACCCGGT----- 144  
 DB 72 ArgGluSerValProSerProLysAspLysAspLeuIleGlnLeuProGlyGlnPro 91  
 QY 145 -----GTCAATCTACTCTGTGATATGTCGACACCTCTCCGAGTCCATACC 192  
 DB 92 SerAspValThrPheLysLysIleTyrglyGlyTyrValAlaValAsnLysProAlaGlyArg 111  
 QY 193 TTC-----TTCGTGTTCTTCGACCGACATACCCAGAACTGCACCTATACATTG 246  
 DB 112 PheLeuTyrTyrPheValGluThrIleLysProGlyAsnThrThrProLeuValIle 131  
 QY 247 TGGTGAATGTGGCCCTGGAAGCATTTTGTATCGTCTCTTCGACAGTTGGCCCT 306  
 DB 132 TrpPheAsnGlyGlyProGlyCysSerSerLeuGlyGlyAlaPheLysGluLeuGlyPro 151  
 QY 307 TGCCATGTCATTCGACTTTTGATGAC---TACATCAACCTCTCACTCGTGAACAGGTG 363  
 DB 152 PheArgValHisSerAspGlyLysThrLeuPheArgAsnProTyrSerTrpAsnAsnGlu 171  
 QY 364 TCCAAATTACATTCTCTCCAGCATTCGGAGTCGGCTTTTCATATAGTATACGGTT 423  
 DB 172 AlaAsnValLeuPheLeuGluThrProValGlyThrGlyPheSerTyrSer----- 188  
 QY 424 GATGGTCCATTAATCCCTGTAACTGCGGGTCTGCGAAATTCGAGCTTTCGAGGATTCAG 483  
 DB 189 -----AsnSerProIleAsnGlyLysGln 196  
 QY 484 GCGCGGTACCAACCATTTGATGCCACTCTGTGATGATACTACCAATCTTTCGCCACAGGCC 543

DB 197 GlyAspLysAlaThr-----AlaGluAsp 204  
 QY 544 GCTGGGAGATCTCTCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGGTCCAG 603  
 DB 205 AsnTyrMetPheLeuValAsnTrpLeuGluArgPheProGluTyrLysGlyArg----- 222  
 QY 604 TCTAGGACTTCAGTCTATGGACGGAGAGCTATGGAGGCGACTAGTGGTCTGCTGCAATCTTC 663  
 DB 223 -----AspIleTyrIleAlaGlyGlnSerTyrAlaGlyHisTyrValProGlnLeuAla 240  
 QY 664 AATCATTTTACGAGCAGAAATGAGAGAATGCCAACGGTAGTGTATGGTGTTCAGCTT 723  
 DB 241 GlnIleLeuHisArgAsn-----AsnGlnThrLeuIle 252  
 QY 724 AATTTCAACTCTCTGGGAATTTAATACGCATCATCGACGAGCGCATCCAGGCCCTTAC 783  
 DB 253 AsnLeuArgGlyIleLeuIleGlyAsnProSerLeuAsnArgGluIleGlnAspPhe 272  
 QY 784 TACCTGGAATTCGCTGTGAACAATACCTACGATGATCAAGCTGTCAACGAGACCGTCTAC 843  
 DB 273 GlyTyrLysPheMetPhe-----SerHisGlyLeuIleSer---GlnGlnGlnMetAsp 289  
 QY 844 AACTCATCATCAAGTTTCCCAACCAATGCCAATGTTGCCAGCATTTGATTTCCACCTGC 903  
 DB 290 AsnTyrAsnLysPhe-----CysThrAsp-----SerAspLeu 300  
 QY 904 AAACAGACAAACCGCACCGCATTAGTACTAGTCTCTGCGCTCTGCGGAGAACCCACACATG 963  
 DB 301 TyrAspTrpAspLysCysHisLeuAlaSerGlnLysIleGluAlaGlnLysThrHisLeu 320  
 QY 964 TCCAGGGACAATGTTGAGGGGCCATACTACGCTTGTGCTGCTGCTGTGTATGATAT 1023  
 DB 321 -----AspIleTyrAsnIle 325  
 QY 1024 CGGCATCCATAT----- 1035  
 DB 326 TyrAlaProLeuCysLeuAsnSerThrLeuSerSerGluProLysLysCysThrThrIle 345  
 QY 1036 -----GATGACCGCACCTCCGCAAGTATTACAAATTTCTGCCAAGGACTCTGTC 1089  
 DB 346 MetLysAlaAspProCysSerGlyAsnTyrLeuLysAlaTyrLeuAsnLysGluVal 365  
 QY 1090 ATGGACGCTATCGGCTCAAC-----ATCACTACACCCAGTCCCAATAT 1134  
 DB 366 GlnGluAlaIleHisAlaAsnThrThrLysIleProTyrGluTrpThrSerCysAsnThr 385  
 QY 1135 GACGTCTACTAGCTTTCAGCAACAGCGGCTTGTCTGCGCCCAACTTCATCGAAGAC 1194  
 DB 386 LysLeuLeuTrpGluTrpAsnGluLysAspArgTyrVal-----SerLeuThrProIle 403  
 QY 1195 CTCGAGGAGATCTTCTCTCTCCCGTGGTCTCTCCCTCATCTATGCGGACCGCATAC 1254  
 DB 404 LeuGlnGluLeuMetGlyLysGlyValArgValMetLeuTyrAsnGlyAspValAspLeu 423  
 QY 1255 ATCTGCACTGGTTCGGGCTCAGGCGGTTCCCTCGCTGCGCACTACTCCCAAGCCGCC 1314  
 DB 424 ValIleProPheThrSerThrLeuAlaValLysThrMetAsnLeuThrValLys 443  
 QY 1315 CAGTTCGGAAGCGGAGGTACAGCGCTGAAAGTCAACGCGCTCGAGTATCGGGAAGCT 1374  
 DB 444 GluTrpArg-----ProTrpPheThrGlyHisValGlyGlyPheThr 458  
 QY 1375 CCGGAGTAT---GGTAATTTCT 1431  
 DB 459 GluAspTyrLysGlyAsnLeuThrPheValThrValLysGlyAlaGlyHisSerValPro 478  
 QY 1432 TACTACGACCGCATCGCT 1476  
 DB 479 ThrAspGlnProIleHisAlaLeuAsnIlePheThrSerPheIle 493  
 RESULT 15  
 ABR38843

ID ABR38843 standard; Protein; 455 AA.

AC ABR38843;

DT 24-APR-2003 (first entry)

DE A. niger carboxypeptidase Y polypeptide #2.

XX Protease; fungal infection; aspergillosis; food; tanning; detergent;  
KW protein solubility; viscosity; taste; texture; nutritional value;  
KW EC3.4.16.5.

OS Aspergillus niger.

XX WO200269623-A2.

PN 06-SEP-2002.

XX 22-FEB-2002; 2002WO-EP01984.

PR 23-FEB-2001; 2001EP-0200657.

PR 23-FEB-2001; 2001EP-0200658.

PR 23-FEB-2001; 2001EP-0200660.

PR 26-FEB-2001; 2001EP-0200706.

PR 26-FEB-2001; 2001EP-0200707.

PR 26-FEB-2001; 2001EP-0200708.

PR 26-FEB-2001; 2001EP-0200719.

PR 28-MAR-2001; 2001EP-0000075.

PR 28-MAR-2001; 2001EP-0000078.

PR 28-MAR-2001; 2001EP-0000080.

PR 28-MAR-2001; 2001EP-0000087.

PR 28-MAR-2001; 2001EP-0000088.

PR 21-MAY-2001; 2001EP-0000156.

PR 21-MAY-2001; 2001EP-0000159.

PR 21-MAY-2001; 2001EP-0000160.

PR 21-MAY-2001; 2001EP-0000162.

PR 21-MAY-2001; 2001EP-0000165.

PR 21-MAY-2001; 2001EP-0000166.

PR 21-MAY-2001; 2001EP-0000168.

PR 21-JUN-2001; 2001EP-0000240.

PR 21-JUN-2001; 2001EP-0000242.

PR 21-JUN-2001; 2001EP-0000244.

PR 21-JUN-2001; 2001EP-0000246.

PR 12-JUL-2001; 2001EP-0000280.

PR 12-JUL-2001; 2001EP-0000285.

PR 30-JUL-2001; 2001EP-0000323.

PR 30-JUL-2001; 2001EP-0000327.

PR 02-AUG-2001; 2001EP-0000341.

PR 02-AUG-2001; 2001EP-0000342.

PR 02-AUG-2001; 2001EP-0000343.

PR 02-AUG-2001; 2001EP-0000344.

PR 09-AUG-2001; 2001EP-0000357.

PR 16-AUG-2001; 2001EP-0000374.

PR 16-AUG-2001; 2001EP-0000377.

PR 20-SEP-2001; 2001EP-0000478.

PR 20-SEP-2001; 2001EP-0000483.

PR 22-OCT-2001; 2001EP-0000552.

PR 22-OCT-2001; 2001EP-0000553.

PR 22-OCT-2001; 2001EP-0000554.

PR 22-OCT-2001; 2001EP-0000556.

PR 22-OCT-2001; 2001EP-0000557.

PR 22-OCT-2001; 2001EP-0000558.

PR 15-NOV-2001; 2001EP-0000464.

PR 21-DEC-2001; 2001EP-0000517.

XX (STAM ) DSM NV.

XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;

PI Kludbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;

PI Maier D, Spreafico F, Folkers U, Kemmer S, Kemmer W, Tan P;

PI Stiebeler J, Albang R;

XX WPI; 2002-723203/78.

DR

DR P-PSDB; ABZ78210, ABZ78267.

XX Novel isolated protease polypeptide useful in laboratory, clinical,  
PT pharmaceutical, chemical, diagnostic, personal care and industrial  
PT applications

XX Claim 13; Page 329-331; 394pp; English.

XX The invention relates to a novel isolated protease polypeptide. A  
CC polypeptide or polynucleotide of the invention is useful for diagnosing a  
CC fungal infection such as aspergillosis, or as a query sequence to perform  
CC a search against public databases. A polypeptide of the invention is  
CC useful in a selected number of industrial or pharmaceutical processes, in  
CC laboratory or clinical processes, in food industry (baking, brewing,  
CC cheese manufacturing, meat tenderizing), in tanning industry and in the  
CC manufacture of biological detergents. A polypeptide may also be useful  
CC for improving protein solubility, extraction yields, viscosity or taste,  
CC texture, nutritional value, minimising of antigenicity or  
CC anti-nutritional factors, colour or functionality as well as processing  
CC aspects like filterability of the proteinaceous raw material. The  
CC sequences shown in ABR38313-ABR38869 represent the A. niger proteases of  
CC the invention.

XX Sequence 455 AA;

SQ Alignment Scores:

Pred. No.:	1-97e-24	Length:	455
Score:	366.50	Matches:	138
Percent Similarity:	40.83%	Conservative:	69
Best Local Similarity:	27.22%	Mismatches:	182
Query Match:	12.62%	Indels:	118
DB:	23	Gaps:	22

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x ABR38843 (1-455)

QY	7	GGAAGTACACGGCGTCCGGTACAGACACACTACCAAGAACCCACCCGGGGTCAAG	66
DB	25	GlyAlaAspProPheThrValPheArgSer-----ProHisSerProAla-----	39
QY	67	ACTCTTACAAACCGCAACAAATGTCACCATCGGTACAGGACCCGGCGAGGGCGTC	126
DB	40	-----PheSerIleArgIleGlnGluLnsn---AspSerIle	51
QY	127	TGCGAGACTACCCCGGGTGTCAAACTCTACTGTGATATGTCGACACCTCTCCGAGTCC	186
DB	52	CysAspAlaArgSer-----ProGlnPheThrGlyTrpLeuAspIleGlyProLys---	68
QY	187	CATACCTTCTCTGGTTCCTCGAAGCCAGACATACCCAGAACTGCATATCATTCATG	246
DB	69	HisLeuPhePheTrpTyrPheGluSerGlnAsnAspProPheHisAspProLeuThrLeu	88
QY	247	TGCTTGAATGGTGGCCCTGGAAGCGATTCTTTGATCGGTCTCTCGAAGAGTTGGCCCT	306
DB	89	TrpMetThrGlyGlyProGlyAspSerSerSerMetIleGlyLeuPheGluGluValGlyPro	108
QY	307	TGCCATGTCAAT-----TCGACTTTTGATGACTACATCAACCCCTCATCTCGTGAACGAG	360
DB	109	CysArgIleAsnGluPheGlyAsnGlyThrAspHis---AsnProTrpAlaTrpThrLys	127
QY	361	GTCTCCAATTACTATTCTCTCCAGCCATTGGAGTGGGCTTTTCATATAGTGATAGC	420
DB	128	AsnSerSerLeuLeuPheValAspGlnProValAspValGlyPheSerTyrIleAsp---	146
QY	421	GTGTATGGGTCCATTAAACCTGTACTGGGGTCTGCGAAATTCGAGCTTCGAGAGTT	480
DB	147	-----GluGlyTyr	149
QY	481	CAGGCCCGGTACCCAAACCATTTGATGCACTCTGTATGATACATACCAATTCGCCGAGAG	540
DB	150	Glu-----LeuProHisAspSerArgGluAlaLaval	160
QY	541	GCCGCTGGGAGATCTGCAAGGATTCCTTAGTGACIACCTAGCTAGCTAGGCTG	600

```
Db      161 AspMethHisArgPheLeuArgLeuPheIleSer-----GluIlePheProHisLys 177
QY      601 CAGTCTAAGGACTTCAGTCTATGACGAGGAGACIATGAGGGACACIATGTCCTCCGCAATC 660
Db      178 GlnPheLeuProValHisLeuSerGlyIleSerTyrAlaGlyArgTyrIleProTyrLeu 197
QY      661 TTCATCATTTTACGAGCAGAAATGAG----- 687
Db      198 AlaThrGlnIleLeuGluGlnAsnGluLeuTyrLysAspSerProArgIleProLeuLys 217
QY      688 -----AGATTTGCCACGGTAGTGTAAATGGTTCAGCTTAATTTCAACTCTCTGGGA 741
Db      218 SerCysLeuValGlyAsnGlyPheMetSerProLysAspAlaThrPheGly-TyrTrpG1 237
QY      742 ATTTATTAAAGGCATCATCGAGCGGATCCAGGCCCTTACTACCCCTGAATTCGCTGTG 801
Db      237 uThrLeuCysThrThrAsnSer-GlyValProSerProIle-Phe----- 251
QY      802 AACAAATACCTACGGTATCAAGGCTGTCAAGGAGACCGCTCTACAATACATGAAGTTTGC 861
Db      252 -----AsnGluThrArgCysAspIleMet-----Ala 260
QY      862 AACCAATGCCAATGTTGCCAGGATTTGATTTCCACCTGCAACACACAAACCCGACC 921
Db      261 AlaAsnMetProHis---CysMetAspLeuTyrAspIleCysIleGlnHis----- 276
QY      922 GCATTAGCTGACTACGCCCTCTGCGCGAGCCACCAACATGTGCAGGACAAATGTTGAG 981
Db      277 -----SerAspProAlaIleCysHisAlaGlnSerValCysTyrAspSerValVal 294
QY      982 GGGCCATPACTACGCCCTTTGCTGGTGGTGGTGTGATGATATTCGGCATCCATATGATGAC 1041
Db      295 GlyLeuMetAlaLysLeuLeuArg-----MetThrThr 306
QY      1042 CCGACTCCGCCA-----AGTTATTACACAAA 1068
Db      307 ValThrAlaProCysGluIleAspGluMetCysTyrIleGluAlaAlaLeuIleGluArg 326
QY      1069 TTCTGCAAGAGGACTCTGATGACGCGCTATCGCGCTCAACATCACTACACCCAGTCC 1128
Db      327 TyrLeuAsnSerProSerValTrpGluAlaLeuSerProGlnGlnValThrGluTyr 346
QY      1129 AAT-----AATGACGCTACTACGCTTTCAGCAACAGCGGACTTTGCTGGGCC 1179
Db      347 LysPheValAlaThrSerValIleAspAlaPheAlaGlnSerAlaAsp----- 362
QY      1180 AACTTCATCGACACCTCGAGGAGATC-----CTTGCTCTCCCGTCCGCTGTCTCC 1230
Db      363 GlyMetValSerSerLysGlnIleAlaPheLeuLeuAlaAsnValAspPheLeu 382
QY      1231 CTCATCTATGCGACGCGGATTACATCTGCAACTGTTGCGCGGTACAGCGGCTTCCCTC 1290
Db      383 AlaTyrGlnGlyAsnLeuLeuAlaCysAsnThrAlaGlyAsnLeuArgIrpAlaAsn 402
QY      1291 GCTCGCAACTACTCCCAAGCGCGCGAGTTCGAGCGCAGGGTACAGCGCCCTGAAAGTC 1350
Db      403 SerLeuSerTrpLysGlyGlnThrGluPheThrAlaLysProLeuLeuProTrpGluIle 422
QY      1351 AAGCGGCTCGAGTATGGGAA---ACTCGCGAGTATGGTAAATTTCTCTTCTACATCGGCTC 1407
Db      423 Gln---ValSerValGlyGluGlyThrAspGluThrSerArgPheAlaPheValThrVal 441
QY      1408 TATGAGCGAGGCCAT 1422
Db      442 AspAsnAlaGlyHis 446
```

**This Page Blank (uspto)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 16, 2003, 18:43:14 ; Search time 192 Seconds  
(without alignments)  
2493.964 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662  
Perfect score: 2904  
Sequence: 1 ctccaggaagtagcacggc.....gcattgcaggttggtgatg 1608

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 556269 seqs, 148893369 residues

Total number of hits satisfying chosen parameters: 1112538

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool.p/US09712338/runat\_16092003\_144327\_14435/app\_query.fasta\_1.1799  
-DB=published.Applications\_AA\_QFWT-fastan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCI=0 -LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09712338 -QCN1.1.15 -runat\_16092003\_144327\_14435  
-NCFU=6 -ICPU=3 -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_AA:

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/BCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARTES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	517.5	17.8	421	9	US-09-420-785A-4	Sequence 4, Appli
2	517.5	17.8	421	9	US-09-901-252-15	Sequence 15, Appl
3	361	12.4	476	14	US-10-084-018-3	Sequence 3, Appli
4	359	12.4	476	10	US-09-909-320-164	Sequence 164, App
5	359	12.4	476	10	US-09-909-088B-164	Sequence 164, App
6	359	12.4	476	10	US-09-905-291A-164	Sequence 164, App
7	359	12.4	476	10	US-09-902-853-164	Sequence 164, App
8	359	12.4	476	10	US-09-907-824-164	Sequence 164, App
9	359	12.4	476	10	US-09-907-841-164	Sequence 164, App
10	359	12.4	476	11	US-09-904-011-164	Sequence 164, App
11	359	12.4	476	11	US-09-906-742-164	Sequence 164, App
12	359	12.4	476	11	US-09-906-838-164	Sequence 164, App
13	359	12.4	476	11	US-09-907-613-164	Sequence 164, App
14	359	12.4	476	11	US-09-907-942-164	Sequence 164, App
15	359	12.4	476	11	US-09-796-753-40	Sequence 40, Appl
16	359	12.4	476	11	US-09-904-859-164	Sequence 164, App
17	359	12.4	476	11	US-09-909-204-164	Sequence 164, App
18	359	12.4	476	11	US-09-904-820-164	Sequence 164, App
19	359	12.4	476	11	US-09-904-786-164	Sequence 164, App
20	359	12.4	476	11	US-09-906-646-164	Sequence 164, App
21	359	12.4	476	11	US-09-906-700-164	Sequence 164, App
22	359	12.4	476	11	US-09-903-786-164	Sequence 164, App
23	359	12.4	476	11	US-09-902-903-164	Sequence 164, App
24	359	12.4	476	11	US-09-903-749A-164	Sequence 164, App
25	359	12.4	476	11	US-09-904-119-164	Sequence 164, App
26	359	12.4	476	11	US-09-904-956-164	Sequence 164, App
27	359	12.4	476	11	US-09-902-736-164	Sequence 164, App
28	359	12.4	476	11	US-09-907-794-164	Sequence 164, App
29	359	12.4	476	11	US-09-903-943-164	Sequence 164, App
30	359	12.4	476	11	US-09-904-462-164	Sequence 164, App
31	359	12.4	476	11	US-09-907-925-164	Sequence 164, App
32	359	12.4	476	11	US-09-902-692-164	Sequence 164, App
33	359	12.4	476	11	US-09-903-520-164	Sequence 164, App
34	359	12.4	476	11	US-09-905-056-164	Sequence 164, App
35	359	12.4	476	11	US-09-909-064-164	Sequence 164, App
36	359	12.4	476	11	US-09-904-553-164	Sequence 164, App
37	359	12.4	476	11	US-09-905-381-164	Sequence 164, App
38	359	12.4	476	11	US-09-905-088-164	Sequence 164, App
39	359	12.4	476	11	US-09-907-575-164	Sequence 164, App
40	359	12.4	476	11	US-09-905-075-164	Sequence 164, App
41	359	12.4	476	11	US-09-902-759-164	Sequence 164, App
42	359	12.4	476	11	US-09-902-634-164	Sequence 164, App
43	359	12.4	476	11	US-09-902-713-164	Sequence 164, App
44	359	12.4	476	11	US-09-907-979-164	Sequence 164, App
45	359	12.4	476	11	US-09-902-615-164	Sequence 164, App

ALIGNMENTS

RESULT 1  
US-09-420-785A-4  
; Sequence 4, Application US/09420785A  
; Patent No. US20010010923A1  
; GENERAL INFORMATION:  
; APPLICANT: MORTENSEN, UFFE  
; APPLICANT: OLESEN, KJELD  
; APPLICANT: STENNICKE, HENNING  
; APPLICANT: SORESENSEN, STEEN B.  
; APPLICANT: BREDDAM, KLAUS  
; TITLE OF INVENTION: MODIFIED CARBOXYPEPTIDASE  
; FILE REFERENCE: 089187/0109  
; CURRENT APPLICATION NUMBER: US/09/420,785A  
; CURRENT FILING DATE: 1999-10-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 421  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-420-785A-4

Alignment Scores: 4.76e-35 Length: 421  
Pred. No.: 4.76e-35 Length: 421

Score: 517.50 Matches: 142  
 Percent Similarity: 43.98% Conservative: 70  
 Best Local Similarity: 29.46% Mismatches: 185  
 Query Match: 17.82% Indels: 85  
 DB: 9 Gaps: 17

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-09-420-785A-4 (1-421)

```

QY 97 CGGTACAAGAACCCGGGCGAGAGCGCTCGGAGACTACCCGGGTGTCAAATCTAC 156
Db 1 LysileLysAspProLysileLeuGlyle-----AspProAsnValThrGlnTyr 17
QY 157 TCTGGATATCGACACCTCTCCCGAG---TCOCATACCTTCTCTGGTTCGAGACC 213
Db 18 ThrGlyTyrLeuAspValGluAspGluAspLysHisPhePheThrPheGluSer 37
QY 214 AGCATAAACCCGAAACGTCACCATATCATATGTGGTTGAATGGTCCCTGGAAGCAT 273
Db 38 ArgAsnAspProAlaLysAspProValleLeuTrpLeuAsnGlyGlyProGlyCysSer 57
QY 274 TCTTTGATCGTCTCTCGAAGAGTGGCGCTTGCCATGTCATTCGATCTTTTGATGAC 333
Db 58 SerLeuThrGlyLeuPhePheGluLeuGlyProSerSerileGlyProAspLeuLysPro 77
QY 334 TACATCAACCTCACTCGTGAACAGAGTCTCCAAATTTACTATTCCTGTCCAGCCATTG 393
Db 78 IleGlyAsnProTyrSerTrpAsnSerAsnAlaThrVallePheLeuAspGlnProVal 97
QY 394 GGAGTCGGCTTTCATATATGATACGGTTGATGGGTCCATTAAACCTGTAAACWGGGTC 453
Db 98 AsnValGlyPheSerTyrSer----- 104
QY 454 GTCGAATTCGAGCTTTCGAGGAGTTCAGGCGGTGACCAACCACTGATGCCACTG 513
Db 105 -----GlySerSerGly----- 108
QY 514 ATCGATATACCAATCTGCGGAGGCGCTTGGGAGATCTCGAAGGATTCCTTAGT 573
Db 109 ValSerAsnThrValAlaAlaGlyLysAspValTyrAsnPheLeuGluLeuPhePheAsp 128
QY 574 GGACTACTAGCTGGACTAGGTGTCAGTCTAAGACTTCAGTCTATGACGAGGAGC 633
Db 129 GlnPheProGluTyrValAsnLysGlyGln-----AspPheHisIleAlaGlyGluSer 146
QY 634 TATGAGGAGGACTATGCTCTGATCTTCAATCATTTTACGAGCAGAAAGAGAGATT 693
Db 147 TyrAlaGlyHisTyrIleProValPheAlaSerGluLeuLeuSerHisLysAspArg--- 165
QY 694 GCCAAGGTAGTGTATATGGTGTTCAGCTTAATTTCACTCTCTGGGAATTAAT----- 747
Db 166 -----AsnPheAsnLeuThrSerValLeuIleGly 175
QY 748 AACGGCATATCGACGAGGCGATCCAGGCCCTTACTACCCCTGATTCGCT----- 798
Db 176 AsnGlyLeuThrAspProLeuThrGlnTyrAsnTyrTyrGluProMetAlaCysGlyGlu 195
QY 799 -----GTGAACAATACCTACGATGATCAAGCTGTCAACGACGACCGCTAC 843
Db 196 GlyGlyGluProSerValLeuProSerGluGluCysSerAlaMetGluAspSerLeu--- 214
QY 844 AACTACATGAAGTTTGGCAACCAATGCCAATGTTGCCAGGATTTGATTCACCTGCG 903
Db 215 -----GluArgCysLeuGlyLeuIleGluSerCys 224
QY 904 AAACAGACAAACCGACCGCATTAGCTACATACGCTCGCGGAGGACCAACCAATG 963
Db 225 TyrAspSerGln-----SerValTrpSerCysValProAlaThrIleTyr 239
QY 964 TCCAGGGAATTTGAGGGGCGCATACGCTTGTGCTGGTGTGCTGTATGATATT 1023
Db 240 CysAsnAsnAlaGlnLeuAlaProTyrGlnArg---ThrGlyArgAsnValTyrAspIle 258
QY 1024 CGGCATCCATATGATGAC-----CCGACTCCGCCCAAGTTATTACACANA 1068

```

```

Db 259 ArgLysAspCysGluGlyGlyAsnLeuCysTyrProThrLeuGlnAsp---IleAspAsp 277
QY 1069 TTTCTGCAAGAGGACTCTGTATGAGCGCTATGGCGTCAACACTCAACTACACCGCTCC 1128
Db 278 TyrLeuAsnGlnAspTyrValLysGluAlaValGlyAlaGluValAspHisTyrGluSer 297
QY 1129 ---AATAATGAGCTTACTACGCTTTCCAGCAACAGGAGGAGCTTTGTCGCGCAACTTC 1185
Db 298 CysAsnPheAspIleAsuAlaGAsnPheLeuPheAlaGlyAspTrpMetLysPro---Tyr 316
QY 1186 ATCGAAGACCTCGAGGAGATCTTCTCTCCCGTGGTCTCCCTCATCTATGCGGAC 1245
Db 317 HisThrAlaValThrAspLeuLeuAsnGlnAspLeuProIleLeuValTyrAlaGlyAsp 336
QY 1246 GCCATATCATCTCAACTGCTGCGCGGTTCAGCGCGTTTCCCTCGCTGCGAAGACTTCC 1305
Db 337 LysAspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrAspValLeuProTrpLys 356
QY 1306 CAAGCCGCCAGTTCGCGAAGC-----GCAGGGTACAGCGCCCTGAAAGTCAACGCC 1356
Db 357 TyrAspGluGluPheAlaSerGlnLysValArgAsnTrpThrAlaSerIleThrAspGlu 376
QY 1357 GTCGAGTATGGGAACTCGGAGTATGTAATTTCTCTTCACTCGGCTATGAGGCA 1416
Db 377 ValAla---GlyGluValLysSerTyrLysHisPheThrTyrLeuArgValPheAsnGly 395
QY 1417 GGCCATGAAGTCCCATCTACAGCCATCCCTCGCTCAATTTTAAACGCGGACTATC 1476
Db 396 GlyHisMetValProPheAspValProGluAsnAlaLeuSerMetValAsnGluTrpIle 415
QY 1477 TTCGGT 1482
Db 416 HisGly 417

```

## RESULT 2

US-09-901-252-15  
 ; Sequence 15, Application US/09901252  
 ; Patent No. US20020026658A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chagale, Clint  
 ; TITLE OF INVENTION: Genes Encoding Sinapoylglucose:Malate Sinapoyltransferase and  
 ; TITLE OF INVENTION: Use  
 ; FILE REFERENCE: N1422-004  
 ; CURRENT APPLICATION NUMBER: US/09/901,252  
 ; CURRENT FILING DATE: 2001-07-09  
 ; PRIOR APPLICATION NUMBER: US 60/216593  
 ; PRIOR FILING DATE: 2000-07-07  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: Patent in version 3.0  
 ; SEQ ID NO 15  
 ; LENGTH: 421  
 ; TYPE: PRT  
 ; ORGANISM: Saccharomyces  
 US-09-901-252-15

Alignment Scores:  
 Pred. No.: 4,76e-35 Length: 421  
 Score: 517.50 Matches: 142  
 Percent Similarity: 43.98% Conservative: 70  
 Best Local Similarity: 29.46% Mismatches: 185  
 Query Match: 17.82% Indels: 85  
 DB: 9 Gaps: 17

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-09-901-252-15 (1-421)

```

QY 97 CGGTACAAGAACCCGGGCGAGAGCGCTCGGAGACTACCCGGGTGTCAAATCTAC 156
Db 1 LysileLysAspProLysileLeuGlyle-----AspProAsnValThrGlnTyr 17
QY 157 TCTGGATATCGACACCTCTCCCGAG---TCOCATACCTTCTCTGGTTCGAGACC 213
Db 18 ThrGlyTyrLeuAspValGluAspGluAspLysHisPhePheThrPheGluSer 37

```



QY	214	AGACATAACCCAGAAATCGCACATACATATGTGTTGAATGGNGCCCTCGAAGCGAT	273
Db	38	ArgAsnAspProAlaLysAspProValIleLeuTrpLeuAsnGlyGlyProGlyCysSer	57
QY	274	TCATTGATCGGCTCTTCGAAAGAGWTGGGCCCTGCCATGTCAAATTCGACTTTTCATGAC	333
Db	58	SerLeuThrGlyLeuPheGluLeuGlyProSerSerIleGlyProAspLeuLysPro	77
QY	334	TACATCAACCCCTACTCTGGAACCCAGTCTCCAAATTTACTATTCCTGTCCACGCATG	393
Db	78	IleGlyAsnProTyrSerTrpAsnSerAsnAlaThrValIlePheLeuAspGlnProVal	97
QY	394	GGAGTCGGCTTTTCATATAGTGATACGGTTGATGGTCCATTAAACCCGTAACTGGGGTC	453
Db	98	AsnValGlyPheSerTyrSer	104
QY	454	GTCGAANAATTCGAGTTTGGCAGGATTCAGGCGCGGTACCAACAACATTCATGCCACTGTG	513
Db	105	-----GlySerSerGly-----	108
QY	514	ATCGATACTACCAATCTTTGCCGACAGAGCCGCTTCGGAGATCCTGCAAGGATTCCTTAGT	573
Db	109	ValSerAsnThrValAlaAlaGlyLysAspValTyrAsnPheLeuGluLeuPhePheAsp	128
QY	574	GGACTACCTAGCTTTGGACTCTAGGGTCGAGTCAAGGACTTCAGTCTATGACACGAGAGC	633
Db	129	GlnPheProGluTyrValAsnLysGlyGln-----AspPheHisIleAlaGlyGluSer	146
QY	634	TATCGAGGGCATATGGCTTCGCAATCTTTCATCATTTTTCACGACGAGAATGAGAGAATT	693
Db	147	TyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeuSerHisLysAspArg--	165
QY	694	GCCAAACGCTAGTGTTAATGGTGTTCACTTAATTTCAACTCTCTCGGAATTAATT-----	747
Db	166	-----AsnPheAsnLeuThrSerValLeuIleGly	175
QY	748	AACGGCATATPCGACGAGCGGATCCAGGCCCTTACTACCCGTGAATTCGCT-----	798
Db	176	AsnGlyLeuThrAspProLeuThrGluTyrAsnTyrTyrGluProMetAlaCysGlyGlu	195
QY	799	-----GTGACACATACCTACGGTATCAAGGTGTCAACGACGACCGCTCTAC	843
Db	196	GlyGlyGluProSerValLeuProSerGluGluCysSerAlaMetGluAspSerLeu--	214
QY	844	AACTACATGAAGTTTGCCAAACCAAAATGCCAAATGGTTGCCAGGATTTGATTTCCACCTGC	903
Db	215	-----GluArgCysLeuGlyLeuIleGluSerCys	224
QY	904	AAACAGACAAACCGCACCGCATAGCTAGCTAGCCCTCTCGCGCGAAGCCACCAACATG	963
Db	225	TyrAspSerGln-----SerValTrpSerCysValProAlaThrIleTyr	239
QY	964	IGCAGGGACAATGTTGAGGGGCCATACTACGCCCTTCTGGTGGTGGTGTGTGTGTGTGT	1023
Db	240	CysAsnAsnAlaGlnLeuAlaProTyrGlnArg---ThrGlyArgAsnValTyrAspIle	258
QY	1024	CGGCATCCCATATATGAC-----CCGACTCGGCCCAAGTTTATACACAAA	1068
Db	259	ArgLysAspCysGluGlyGlyAsnLeuCysTyrProThrLeuGlnAsp---IleAspAsp	277
QY	1069	TTTCTGCGCAAGGACTCTGTGATGGAGCTATCGGCGTCAACATCAACTACACCCAGTCC	1128
Db	278	TyrLeuAsnGlnAspTyrValLysGluAlaValGlyAlaGluValAspHisTyrGluSer	297
QY	1129	---AATTAATCACGCTCTACTACGCTTTTCCACGAACAGCGGCACTTTGTCTGGCCCCAATTC	1185
Db	298	CysAsnPheAspIleAsnArgAsnPheLeuPheAlaGlyAspTrpMetLysPro---Tyr	316
QY	1186	ATCGAAGACCTCGAGGAGATCCTTGCTCTCCCGTGGTGTCTCCCTCATCTATGGCGAC	1245
Db	317	HisThrAlaValThrAspLeuLeuAsnGlnAspLeuProIleLeuValTyrAlaGlyAsp	336

```

QY      1246 GCCGATTACATCTGCACAAGTGTTCGGCGGTCCAGGCCGTTCCCTCGTGCGCAACTACTCC 1303
Db      |||:::|||||:::||| ||| ::||| ::|
337 LysAspPheIleCysAsnTrpLeuGlyAsnLysAlairThrAspValLeuProTrpLys 356
QY      1306 CAAGCGCCCGAGTTCGCCGAAGC-----GCAGGGTACAGCCCCCTGAAAGTCAACGGC 1356
Db      ::||| ||| ::||| ::||| ::||| ::|||
357 TyrAspGluGluPheAlaserEglnLysValArgAsnTrpThrAlaSerIleThrAspGlu 376
QY      1357 CTCGAGTAGTGGGAACATCGCGAGTATGTAATTCCTCCCTCACICGCGTCAAGAGGCA 1416
Db      ||| ||||| ::| ||| ::|||::||| ::|||::|||
377 ValAla---GlyGluValLysSerTyrrLyshisPheThrTyrrLeuArgValPheAsnGly 395
QY      1417 GGCATGAAGTCCCATACTACACGCCCATCGCTCCCTCGCAATGTGTTAACCGGACTATC 1476
Db      ||||| ||||||::| ||| ::||| ::||| ::||| |||
396 GlyHisMetValProPheAspValProGluAsnAlaLeuSerMetValAsnGluTrpIle 415
QY      1477 TTCGGT 1482
Db      |||
416 HisGly 417

RESULT 3
US-10-084-018-3
; Sequence 3, Application US/10084018
; Publication No. US20020160499A1
; GENERAL INFORMATION:
; APPLICANT: Handman, Olga
;              Hawkins, Phillip R.
;              Hillman, Jennifer L.
;              Lal, Preeti
;              Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN SERINE
;                   CARBOXYPEPTIDASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,018
; FILING DATE: 25-Feb-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,689A
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/828,488
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0241 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 476 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: MMLR3DT01
; CLONE: 566993
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-084-018-3

Alignment Scores:
```

Pred. No.:	8,72e-22	Length:	476
Score:	361.00	Matches:	130
Percent Similarity:	39.51%	Conservative:	62
Best Local Similarity:	26.75%	Mismatches:	178
Query Match:	12.43%	Indels:	116
DB:	14	Gaps:	21

  

US-09-712-338-1_COPY_55_1662 (1-1608) x US-10-084-018-3 (1-476)	
QY 139	CGGGT-----GTCAAATCTACTCTGGATAT-----GTGCACACCTCTCCGAGTC 186
DB	:      :      :      :      :      :      :      :      :
QY 167	ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86
QY 187	CATACCTTCCTTCGGTCTTCGAAAGCAGACATPACCCAGAACTGCACCTATPACATG 246
DB	:::       :      :      :      :      :      :      :      :
QY 247	TGGTTGAATGGTGGCCCTGGAAGCGATTCTTCATCGGTCTCTTCGAAAGATGGCCCT 306
DB	
QY 307	TGCCATGTCAAATCGACTTTTGATGACTACATCAACCTCATCTGCGTGGACGAGGTCCTC 366
DB	
QY 367	AAATTACTATTCCTGTCGCCAATGGAGTCGGCTTTCATATAGTCATAGCGTTGAT 426
DB	:      :      :      :      :      :      :      :      :
QY 427	GGGTCCATTAACCTGTAACTGGGGTCGTCGAAATTCGAGCTTCGAGGATCAGGC 486
DB	
QY 487	CGGTACCCACCAATGTATGACCTCTGATCGACTACTACCAATCTGCCGAGAGCGGCT 546
DB	:      :      :      :      :      :      :      :      :
QY 547	TGGGAGATPCNGAAGATTCCTTAGTGACTACCTAGCTTGGACTTAGGGTGCAGTCT 606
DB	:      :      :      :      :      :      :      :      :
QY 607	ANGACTTCAGTCTATGAGCGAGAGCTATGGAGGCACTATGCTCTGCATCTCTCAAT 666
DB	:      :      :      :      :      :      :      :      :
QY 667	CATTTTACGACGAGATGAGAGAAATGCCAACGGTAGTGTAAATGGTGTACGCTTAAT 726
DB	:      :      :      :      :      :      :      :      :
QY 727	TCRACTCTCWWGAATATTAAACGGATCATCGACGAGGATCCAGGCCCCCTACTAC 786
DB	:      :      :      :      :      :      :      :      :
QY 787	CCTGAATTCGTGTGAACAATACTACGGTATCAAGGCTGTCAACGACGCGCTACAAC 846
DB	:      :      :      :      :      :      :      :      :
QY 847	TACATGAAGTTGCCAACCAATGCCAAATGGTTCCGAGGATTTGATTTCCACTGCAAA 906
DB	:      :      :      :      :      :      :      :      :
QY 907	CAGACAAAC-----CGCACCGCATAGCTACGCCCCCTCGCGGAGCCACC 957
DB	:      :      :      :      :      :      :      :      :
QY 958	AAC-----ATGTGAGGGAACAATGTTGAGGG-----CCATACTACGGCTTCGCGGT 1005
DB	:      :      :      :      :      :      :      :      :
QY 1006	CGTGTGTGATGATATTTCGGCATCCATATGATGACCGGACTCGCCCAAGTTATTACAAC 1065
DB	:      :      :      :      :      :      :      :      :
QY 1065	ArqCysThrGluProGluAspGlnLeuTyrTyrVal 330

; PRIOR APPLICATION NUMBER: US 60/143,048  
 ; PRIOR FILING DATE: 1999-07-07  
 ; PRIOR APPLICATION NUMBER: US 60/145,698  
 ; PRIOR FILING DATE: 1999-07-26  
 ; PRIOR APPLICATION NUMBER: US 60/146,222  
 ; PRIOR FILING DATE: 1999-07-28  
 ; PRIOR APPLICATION NUMBER: PCT/US99/20594  
 ; PRIOR FILING DATE: 1999-09-08  
 ; PRIOR APPLICATION NUMBER: PCT/US99/20944  
 ; PRIOR FILING DATE: 1999-09-13  
 ; PRIOR APPLICATION NUMBER: PCT/US99/21090  
 ; PRIOR FILING DATE: 1999-09-15  
 ; PRIOR APPLICATION NUMBER: PCT/US99/21547  
 ; PRIOR FILING DATE: 1999-09-15  
 ; PRIOR APPLICATION NUMBER: PCT/US99/23089  
 ; PRIOR FILING DATE: 1999-10-05  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28214  
 ; PRIOR FILING DATE: 1999-11-29  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28313  
 ; PRIOR FILING DATE: 1999-11-30  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28564  
 ; PRIOR FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28565  
 ; PRIOR FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: PCT/US99/30095  
 ; PRIOR FILING DATE: 1999-12-16  
 ; PRIOR APPLICATION NUMBER: PCT/US99/30911  
 ; PRIOR FILING DATE: 1999-12-20  
 ; PRIOR APPLICATION NUMBER: PCT/US99/30999  
 ; PRIOR FILING DATE: 1999-12-20  
 ; PRIOR APPLICATION NUMBER: PCT/US00/00219  
 ; PRIOR FILING DATE: 2000-01-05  
 ; NUMBER OF SEQ ID NOS: 423  
 ; SEQ ID NO 164  
 ; LENGTH: 476  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-909-320-164

## Alignment Scores:

Pred. No.: 1-29e-21 Length: 476  
 Score: 359.00 Matches: 130  
 Percent Similarity: 39.51% Conservative: 62  
 Best Local Similarity: 26.75% Mismatches: 178  
 Query Match: 12.36% Indels: 116  
 DB: 10 Gaps: 21

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-09-909-320-164 (1-476)

QY 139 CCGGGT-----GTCAATCCCTACTGTGGATAT-----GTCCAGACCTCTCCCGAGTCC 186  
 Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86  
 QY 187 CATACCTTCTCTGCTCTTCAAGCCAGACATACCCAGAACTGCACCTATACATTG 246  
 Db 87 AsnLeuPheThrPheProAlaGlnLeuGlnPheProGluAspAlaProValValLeu 106  
 QY 247 TGGTGAATGTGGCCCTGGAAGCGATTCTTTGATCGGTCTCTCGAAGAGTTGGCCCT 306  
 Db 107 TrpLeuGlnGlyGlyProGlyGlySerSerMetPheGlyLeuPheValGluHisGlyPro 126  
 QY 307 TCCCATGTCAATTCGACCTTTTGTGACTACATCAACCCCTCACTCGTGGAAACGAGTCTCC 366  
 Db 127 TyrValValThrSerAsnMetThrLeuArgAspArgAspPheProTyrThrThrLeu 146  
 QY 367 ATTTTACTATCTCTGCCCGCATGGAGTCGGCTTTTATATAGTATACGGTTGAT 426  
 Db 147 SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspThrHis 166  
 QY 427 GGGTCCATTAAACCCGTAACTGGGGTCTCGTCAAAATTCGAGTTTGCAGGAGTTCAGGCG 486  
 Db 167 Gly-----GGTGGGAT 1488

QY 487 CGGTACCCCAACCATGTATGCCACCTCTGTATCGTACTACCAATCTTGGCGAGAGCCCT 546  
 Db 168 -----TyrAlaValAsnGluAspAspValAlaArgAspLeuTyr 180  
 QY 547 TGGGAGATCTGCAAGGATTCCTTAGTGGACTACCTAGCTGGACTCTAGGCTCAGTCT 606  
 Db 181 SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn 195  
 QY 607 AAGACTTCACTATGTCAGCGAGAGCTATGAGGCGACTATGTCCTGCTGCTTCTCAAT 666  
 Db 196 AsnAspPheIyrValThrGlySerTyrAlaGlyTyrValProAlaIleAlaHis 215  
 QY 667 CATTTTTCAGGACAGATGAGAAATGCCAACGGTAGTGTAAAGTGTTCACCTTAAT 726  
 Db 216 LeuIleHisSerLeuAsn-----ProValArgIleValLysIleAsn 229  
 QY 727 TTCACCTCTCTGGGAATATTAAAGGCATCATCAGGAGGAGATCCAGGCCCTTACTAC 786  
 Db 230 LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyGlyTyr 249  
 QY 787 CCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCACGAGAGACCGTACAAC 846  
 Db 250 AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGluLysGlnLysLys 265  
 QY 847 TACATGAAGTTTGCACCAATGCCAAATGGTTGCCAGGATTTGATTTCCACCTGAAA 906  
 Db 266 Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg 279  
 QY 907 CAGACAAC-----CGCACCGCAATTAGCTACTACGCGCTCTCGCGAGAGCCACC 957  
 Db 280 LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr 299  
 QY 958 AAC-----ATGTGAGGGACAATGTTGAGGG-----CCATACTACGCTTTCTGCTGT 1005  
 Db 300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu--- 318  
 QY 1006 CGTGGTCTGTATGATATTCGCATCATATGATACCGGACTCCGCGCAAGTATTACAAC 1065  
 Db 319 -----ArgCysThrGluProGluAspGlnLeuTyrTyrVal 330  
 QY 1066 AAATTTCTGCAAGGACTCTGTGTCAGCGCTATCGCGCTCAACACTACACCCAG 1125  
 Db 331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349  
 QY 1126 TCCAAATAGTACGCTAC-----TACGCTTTCCAGCAACAGCGGACTTTGCTGCGCC 1179  
 Db 350 AsnAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLysPro 369  
 QY 1180 AACTTTCATCAAGACCTCGAGGAGATCCTTGTCTCCCGTGGTGTCTCCCTCATCTAT 1239  
 Db 370 TrpLeuThrGluIleMetAsnAsn-----TyrLysValLeuIleTyrAsn 384  
 QY 1240 GCGAGCGCGATTATCATC-----TGCAC 1263  
 Db 385 GlyGlnLeuAspIleValAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp 404  
 QY 1264 TGGTTCGGCGGTGAGCCGCTTTCCTCGCTGCGAAGTCTCCCAAGCCGCCAG----- 1317  
 Db 405 TrpLysGlySerGln-----GluTyrLysLysAlaGluLysLysVal 418  
 QY 1318 -----TTCGGRAGC-----GCAGGTACAGCCCTCGAAGTCAACGGC 1356  
 Db 419 TrpLysIlePheLysSerAspSerGluValAlaGlyTyr----- 431  
 QY 1357 GTCAGTATGGGAAACTCGCGAGTATGGTAATTTCTCCTTCTCCTGCTCTATGAGCA 1416  
 Db 432 -----IleArgGlnAlaGlyAspPheHisGlnValIleArgGlyGly 446  
 QY 1417 GCCCATAGTCCCATCTACTACAGCCCATCGCTCCCTGCAATTTGTTTAAACCGGATATC 1476  
 Db 447 GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPhe 466  
 QY 1477 TTC-----GGTGGGAT 1488



```
Db 300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu--- 318
QY 1006 CGTGTGTATGAIAITCGGCATCATATGATGACCGGACTCGCGCAAGTTATACAA 1065
Db 319 -----ArgCysThrGluProGluAspGlnLeuTyrTyrVal 330
QY 1066 AANTTCTGCAGAGGACTGTGATGAGCGCTATCGCGGTCAACATCAACATACACCCAG 1125
Db 331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349
QY 1126 TCCAAATAACAGCTCTAC-----TAGCCTTCCAGCAACAGCGGACTTTGTCGCC 1179
Db 350 AsuAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLysPro 369
QY 1180 AACTTCAATGCAAGACGCGAGGAGATCCITGCTCCCGTCCCGTGCCTCCCTCATCTAT 1239
Db 370 TrpLeuThrGluIleMetAsnAsn-----TyrLysValLeuIleTyrAsn 384
QY 1240 GCGACGCGCATATACATC-----TGCAAC 1263
Db 385 GlyGlnLeuAspIleIleValAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp 404
QY 1264 TGCTTGGCGGTGAGCGCGTTCCTCGCTGCGAATCTCCCAAGCGCGCCAG----- 1317
Db 405 TrpLysGlySerGln-----GluTyrLysLysAlaGluLysLysVal 418
QY 1318 -----TTCGGAAGC-----GCAGGTACACGCCCTCGAAGTCAACGGC 1356
Db 419 TrpLysIlePheLysSerAspSerGluValAlaGlyTyr----- 431
QY 1357 GTGAGTATGGGAAACTCGGAGTATGTTATTCCTTCATCGCGTCTATGAGGCA 1416
Db 432 -----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlyGly 446
QY 1417 GGCCATGAATCCCATCTACAGCCGCGCTCCCTCGAATGTTTACCGGATATC 1476
Db 447 GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle 466
QY 1477 TTC-----GGTGGGAT 1488
Db 467 IyrGlyLysGlyTrpAsp 472
```

RESULT 6

```
US-09-905-291A-164
; Sequence 164, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth. J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,291A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20344
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRI
; ORGANISM: Homo sapiens
US-09-905-291A-164

Alignment Scores:
Pred. No.: 1,29e-21 Length: 476
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.36% Indels: 116
DB: 10 Gaps: 21

US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-905-291A-164 (1-476)
QY 139 CCGGGT-----GTCAATCCCTACTCTGATAT-----GTGACACCTCTCCCGAGTC 186
Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86
QY 187 CATACCTCTCTGTTCTTCGAAGCCAGACATACCCAGAACTCCACATATCATTCATG 246
Db 87 AsnLeuPhePheTrpPheProAlaGlnIleGlnProGluAspAlaProValValLeu 106
QY 247 TGGTTGAATGTGGCCCTGGAAGGATCTTTGATCGTCTCTTCCAGAGTGGGCCCT 306
Db 107 TrpLeuGlnGlyProGlySerMetPheGlyLeuPheValGluHisGlyPro 126
QY 307 TGCCATGTCAATTCGACTTTTGATGACTACATCAACCCCTCATCTCGTGGACGAGTCTCC 366
Db 127 TyrValValThrSerAsnMetThrLeuArgAspArgAspPheProTrpThrThrLeu 146
QY 367 AATTTACTATTCCTGCCAGCCATTGGGAGTCGGCTTTTTCATATAGTATACGGTGTAT 426
```

[illegible]

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.

APPLICANT: ASHKENAZI, AVI

APPLICANT: Botstein, David

APPLICANT: Desnovers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Filvaroff, Ellen

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerber, Hanspeter

APPLICANT: Gertsitsen, Mary E.

APPLICANT: Goddard, A.

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth, J.

APPLICANT: Kijavlin, Ivar J.

APPLICANT: Mather, Jennie P.

APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: 10466-14  
CURRENT APPLICATION NUMBER: US/09/907,824  
CURRENT FILING DATE: 2001-07-17  
PRIOR APPLICATION NUMBER: 09/665,350  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: PCT/US99/28214  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30  
PRIOR APPLICATION NUMBER: PCT/US99/28564  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/28565  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30095  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: PCT/US99/30911  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US99/30999  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US00/00219  
PRIOR FILING DATE: 2000-01-05  
NUMBER OF SEQ ID NOS: 423  
SEQ ID NO 164  
LENGTH: 476  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-09-907-824-164

Alignment Scores:  
Pred. No.: 1.29e-21  
Score: 359.00  
Matches: 476  
Conservative: 62  
Percent Similarity: 39.51%  
Best Local Similarity: 26.75%  
Query Match: 12.36%  
DB: 10  
Gaps: 21

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-09-907-824-164 (1-476)

Qy 139 CCGGGT-----GTCAATCTCTCTGGATAT-----GTCCAGACCTCTCCGAGTCC 186  
Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86  
Qy 187 CATACCTCTCTGGTCTTCGAGCCAGACATACCGAGAACTGCACCATCATCACATTG 246  
Db 87 AsnLeuPhePhePhePhePhePhePhePhePhePhePhePhePhePhePhePhePhePhe 106

QY 247 TGGTTGAATGGTGGCCCTCGAAGCGGATCTTTTCATCGGTCTCTTCAAGAGTTGGCCCT 306  
Db 107 TrpLeuGlnGlyGlyProGlySerMetPheGlyLeuValGlnHisGlyPro 126  
QY 307 TGGCATGTCAATTCGACCTTTTGATGACTACATCAACCTCTACTCGTGGACGAGGTCTCC 366  
Db 127 TyrValValThrSerAsnMetThrLeuArgAspArgAspPheProIleThrThrLeu 146  
QY 367 AATTTACTATTCCTGTCGCCAGCAATTCGGAGTGGCGCTTTTCATATAGTATAGATCGTGTAT 426  
Db 147 SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspThrHis 166  
QY 427 GGGTCCATTAACCTCTAATCIGGGTCTGCGAATAATTCGAGCTTCGACGAGTTCAGGGC 486  
Db 167 Gly----- 167  
QY 487 CGGTACCCACCAATTGATGCGACTCTGATCGATACCTACCAATCTTCCGACGAGCGGCT 546  
Db 168 -----TyrAlaValAsnGluAspValAlaArgAspLeuTyr 180  
QY 547 TGGGAGATCCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGAGCTCTAGGGTGCAGTCT 606  
Db 181 SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn 195  
QY 607 AAGGACTTCAGTCTATGGACGAGAGCCTATGGGGGACACTATGTCCTGCTGCAATTCCTCAAT 666  
Db 196 AsnAspPheTyrValThrGlyGluSerTyrAlaGlyLysTyrValProAlaIleAlaHis 215  
QY 667 CATTTTACGAGCAGAAATGAGAGAAATGCCAAGGTAGTGTATGTTGTTGCTGCTGCTAAT 726  
Db 216 LeuIleHisSerLeuAsn-----ProValArgGluValLysIleAsn 229  
QY 727 TTCAACTCTCTGGGAATTTAATTAAGCGCATCATGACGAGCGGATCCAGGCCCTTACTATAC 786  
Db 230 LeuAsnGlyIleAlaIleGlyAspGlyIleSerAspProGluSerIleIleGlyGlyTyr 249  
QY 787 CCTGAATTCGCTGTGAACATACCTACGGTATCAAGGCTGTCAACGAGACGCTCTACAAC 846  
Db 250 AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGluLysGlnLysLys 265  
QY 847 TACATGAAGTTTGCCAAACCAATGCCAAGTGGTCCAGGATTTGATTTCCACCTGCAAA 906  
Db 266 Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg 279  
QY 907 CAGACAAAC-----CCGACCGCAATTAGCTAGCTACGCCCTCTCGGCCGAGACCCACC 957  
Db 280 LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr 299  
QY 958 AAC-----ATGTCGACGGGCAATCTTCAGGG-----CCATACTAGCCCTTTCGTGTGT 1005  
Db 300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu--- 318  
QY 1006 CGTGGTGTGTATGATATTCGGCATCCATATGATGACCGGACTCCGCCCAAGTATTATACAAC 1065  
Db 319 -----ArgCysThrGluProGluAspGlnLeuTyrVal 330  
QY 1066 AATTTCTGCGAAAGGACTCTGTATGAGCCCTATCGCGCTCAACATCACTACACCCAG 1125  
Db 331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349  
QY 1126 TCCAATATGACGCTAC-----TACGCTTTCAGCAACACAGCGGACTTTGCTGGGCC 1179  
Db 350 AsnAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLysPro 369  
QY 1180 AACTTCATCGAAGACCTCGAGGAGATCCTTCTCTCCCGTCCGCTGCTCCCTCATCAT 1239  
Db 370 TrpLeuThrGluIleMetAsnAsn-----TyrLysValLeuIleTyrAsn 384  
QY 1240 GCGACGCGGATATCATC-----TGCATC-----TGCATC-----TGCATC 1263  
Db 385 GlyGlnLeuAspIleIleValAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp 404  
QY 1264 TGGTTCGCGGGTCAGCGCGTTTCCTCTCGCTGCGAACTACTCCCAAGCGGCCAG----- 1317



```

Db      405  TrpLysGlySerGln-----GluTyrLysLysAlaGluLysLysVal 418
QY      1318  -----TTCCGAAG-----CGAGGTACAGCCCTCGAAAGTCAACGGC 1356
Db      419  TrpLysIlePheLysSerAspSerGluValAlaGlyTyr----- 431
QY      1357  GTCAGATATGGGAAACTCGCAGTAGTGGTAATTCTCTCACTCGCGTCTAIGAGCCA 1416
Db      432  -----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlyGly 446
QY      1417  GCCCATGAATGCCATACCTACCCAGCCATCGCCCTCGCAATGCTTTAACGGGACTATC 1476
Db      447  GlyHisIleLeuPzTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle 466
QY      1477  TTC-----GGTGGGAT 1488
Db      467  TyrGlyLysGlyTyrPasp 472

```

## RESULT 9

```

US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-712-338-1_COPY_55_1662 (1-476)
; Sequence 164, Application US/0907841
; Publication No. US20020198366A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,841
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214

```

```

; PRIOR FILING DATE: 1999-11-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-712-338-1_COPY_55_1662 (1-476)
Alignment Scores:
Pred. No.: 1,29e-21 Length: 476
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.36% Indels: 116
DB: 10 Gaps: 21

```

```

QY      139  CCGGGT-----GTCAAATCCTACTCTGGATAT-----GTGCAACCTCTCCCGAGTCC 186
Db      67  ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86
QY      187  CATACCTTCTCTGTTCTTCGAGCCAGACAIACCAGAACTGCACCTATCACATTG 246
Db      87  AsnLeuPhePheTyrPhePheProAlaGlnIleGlnProGluAspAlaProValLeu 106
QY      247  TGGTTGAATGGTGGCCCTGGAAGGATTCITTTGATCGGTCTCTTCGAAAGAGTTGGGCC 306
Db      107  TrpLeuGlnGlyGlyProGlyGlySerSerMetPheGlyLeuPheValGluHisGlyPro 126
QY      307  TGGCATGTCAATTCGACTTTTGTATGACTACATCAACCTCTACTCGTGAAGAGGTCTCC 366
Db      127  TyrValValThrSerAsnMetThrLeuArgAspArgAspPheProTyrThrThrLeu 146
QY      367  AATTACTATTTCCTCCAGCCCATTCGGAGTCGGCTTTCATATAGTATGATACGGTTGAT 426
Db      147  SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspAspHis 166
QY      427  GGGTCCATTAAACCTGTAACTGGGGTCTGCGAAATTCGAGCTTTCGAGGAGTTCAGGGC 486
Db      167  Gly----- 167
QY      487  CGGTACCCCAACCATTCATGTCCTGATCACTACTACCAATCTGCGAGAGCGCGCT 546
Db      168  -----TyrAlaValAsnGluAspValAlaAlaArgAspLeuTyr 180
QY      547  TGGGAGATCCTGCAAGGATTCCTTAGTGGACTACTAGCTTGGACTCTAGGGTCACTCT 606
Db      181  SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn 195
QY      607  AAGGACTTCAGTCTATGACGAGAGCTATGAGGAGCAGTATGCTCTGCTGATCTTCAAT 666
Db      196  AsnAspPheTyrValThrGlyGlySerTyrAlaGlyLysTyrValProAlaIleAlaHis 215
QY      667  CATTTTACGACGAGATGACAAATTCGCAAGGTAGTCTTATGCTGTGTACGTTAAT 726
Db      216  LeuIleHisSerLeuAsn-----ProValArgGluValLysIleAsn 229
QY      727  TTCAACTCTCTGGGAATTAATAACGGCATCATCGACGAGCGGATCCAGGCCCTTACTAC 786
Db      230  LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyGlyTyr 249
QY      787  CTTGAATTCGCTCTGAACAATACCTACGTTATCAAGGCTGTCACAGGAGACCGCTACAAC 846
Db      250  AlaGluPhe-----LeuTyrGlnIleGlyLeuAspGluLysGlnLysLys 265
QY      847  TACATGAAGTTTCCCAACCAATGCCAATGGTTGCCAGGATTTGATTTCCACTGCACA 906
Db      266  Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg 279
QY      907  CACACAAAC-----CGCACCGCATTAGCTACGCTCTGCGCGGAGCAAC 957

```

Db 280 LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr 299  
QY 958 AAC -----ATGTGAGGACAAATGTTAGGG-----CCATCTACCGCTTGTGTT 1005  
Db 300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTrpAsnPheLeu--- 318  
QY 1006 CGTGGTGTCTATCATATCGGCATCCATATGATGACCGACTCCGCGCAAGTTATTACAAC 1065  
Db 319 -----ArgCysThrGluProGluAspGlnLeuIleTyrVal 330  
QY 1066 AAATTTCTGCAAGAACACTGTGTCATGAGACGTCATCGCGTCAACATCAACATCACCCAG 1125  
Db 331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349  
QY 1126 TCCAAATAATCAGCTCTAC-----TAGCGTTTCCACGACACGCGGACATTGTGCGCC 1179  
Db 350 AsnAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLysPro 369  
QY 1180 AACTTTCATCAAGACCTCGAGGAGATCTTCTCTCCCGTGGTGTCTCCCTCACTAT 1239  
Db 370 TrpLeuThrGluIleMetAsnAsn-----TyrLysValLeuIleTyrAsn 384  
QY 1240 GCGGACGCGATTCATC-----TGGCAAC 1263  
Db 385 GlyGlnLeuAspIleIleValAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp 404  
QY 1264 TGGTTGCGGGTCCAGCGGTTTCCCTCGCTCGCAACTACTCCCAAGCGCGCCAG----- 1317  
Db 405 TrpLysGlySerGln-----GluTyrLysLysAlaGluLysLysVal 418  
QY 1318 -----TTCGGAAC-----GCAGGGTACACGCGCCCTGAAAGTCAACGGC 1356  
Db 419 TrpLysIlePheLysSerAspSerGluValAlaGlyTyr----- 431  
QY 1357 GTCGAGTATGGGAAACTCCGAGATATGGAATTTCTCCTTCACTCCGCTCAAGGCA 1416  
Db 432 -----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlyGly 446  
QY 1417 GGCATGAGTCCCATACTACCGCCATCGCTCCCTGCAATTTGTTTAAACGACATC 1476  
Db 447 GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle 466  
QY 1477 TTC-----GGTTGGGAT 1488  
Db 467 TyrGlyLysGlyTrpAsp 472

RESULT 10  
US-09-904-011-164  
; Sequence 164, Application US/09904011  
; Publication No. US20030003530A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Mather, Jeanie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/904,011  
; CURRENT FILING DATE: 2001-07-11  
; PRIOR APPLICATION NUMBER: 09/665,350  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: PCT/US00/04414  
; PRIOR FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: US 60/143,048  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 60/145,698  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: US 60/146,222  
; PRIOR FILING DATE: 1999-07-28  
; PRIOR APPLICATION NUMBER: PCT/US99/20594  
; PRIOR FILING DATE: 1999-09-08  
; PRIOR APPLICATION NUMBER: PCT/US99/20944  
; PRIOR FILING DATE: 1999-09-13  
; PRIOR APPLICATION NUMBER: PCT/US99/21090  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/21547  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/23089  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: PCT/US99/28214  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: PCT/US99/28313  
; PRIOR FILING DATE: 1999-11-30  
; PRIOR APPLICATION NUMBER: PCT/US99/28564  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/28565  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/30095  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: PCT/US99/30911  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US99/30999  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US00/00219  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 423  
; SEQ ID NO 164  
; LENGTH: 476  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-09-904-011-164  
Alignment Scores:  
Pred. No.: 1,29e-21 Length: 476  
Score: 359.00 Matches: 130  
Percent Similarity: 39.51% Conservative: 62  
Best local Similarity: 26.75% Mismatches: 178  
Query Match: 12.36% Indels: 116  
DB: 21 Gaps:  
US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-09-904-011-164 (1-476)  
QY 139 CCGGT-----GTCAATCTACTCTGATAT-----GTGACACCTCTCCGAGTCC 186  
Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86  
QY 187 CATACCTTCTTCTGGTCTTCCAGCCAGACATAACCCAGAACCTGCACCTATCATG 246  
Db 87 AsnLeuPhePheTrpPheProAlaGlnIleGlnProGluAspAlaProValValLeu 106  
QY 247 TGGTTGAATGGTGGCCCTGGAGCGATTTGTATCGGTCTCTTTCGAGAGTTGGCCCT 306  
Db 107 TrpLeuGlnGlyGlyProGlyGlySerSerMetPheGlyLeuPheValGluHisGlyPro 126



; PRIOR FILING DATE: 1999-11-30  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28564  
 ; PRIOR FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28565  
 ; PRIOR FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: PCT/US99/30095  
 ; PRIOR FILING DATE: 1999-12-16  
 ; PRIOR APPLICATION NUMBER: PCT/US99/30911  
 ; PRIOR FILING DATE: 1999-12-20  
 ; PRIOR APPLICATION NUMBER: PCT/US99/30999  
 ; PRIOR FILING DATE: 1999-12-20  
 ; PRIOR APPLICATION NUMBER: PCT/US00/00219  
 ; PRIOR FILING DATE: 2000-01-05  
 ; NUMBER OF SEQ ID NOS: 423  
 ; SEQ ID NO 164  
 ; LENGTH: 476  
 ; TYPE: PRT  
 ; ORGANISM: Homo Sapien  
 US-09-906-742-164

## Alignment Scores:

Pred. No.: 1,29e-21 Length: 476  
 Score: 359.00 Matches: 130  
 Percent Similarity: 39.51% Conservative: 62  
 Best Local Similarity: 26.75% Mismatches: 178  
 Query Match: 12.36% Indels: 116  
 DB: 11 Gaps: 21

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-09-906-742-164 (1-476)

QY 139 CCGGCT-----GTCAAACTCTACTCTGGATAT-----GTCCAGCACCTCTCCCGAGTCC 186  
 Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86  
 QY 187 CAPACCTCTCTGGTCTTCGAGCCAGACATACCCAGAACTGCACCTATCACATTS 246  
 Db 87 AsnLeuPheThrPheProAlaGlnProGluAspAlaProValLeu 106  
 QY 247 TGGTGAATGGTCCCTGGAAGCATCTTTGATCGTCTCTCGAAGAGTTGGGCCT 306  
 Db 107 TriPleuGlnGlyProGlyGlySerSerMetPheGlyLeuPheValGluHisGlyPro 126  
 QY 307 TGCCATGTCAATTCGACTTTTGATGACTACATCAACCTCCTCGTGGAGAGGTCTCC 366  
 Db 127 TyrValValThrSerAsnMetThrLeuArgAspPheProTyrThrThrLeu 146  
 QY 367 AATTACTATCTGCCCCAGCATGGGAGTGGCTTTTCATATAGTATGATCGATGAT 426  
 Db 147 SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspThrHis 166  
 QY 427 GGGTCCATTAACTGTAACCTGGGTGCTCGAAATTCGAGCTTTGCAGGAGTTCAGGGC 486  
 Db 167 Gly----- 167  
 QY 487 CGGTACCAACCAATTGATGCCACTTCGATCATCTACCAATCTGCGAGAGCGGCT 546  
 Db 168 -----TyrAlaValAsnGluAspValAlaArgAspLeuTyr 180  
 QY 547 TGGGAGATCTCGAAGGATTCCTTAGTGGACTAGCTTGGACTTAGCGTCCAGTCT 606  
 Db 181 SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn 195  
 QY 607 AAGCACTTCAGTCTATGACGAGGAGGATGATGGAGGCACTATGGTCTCTGCAATCTCAAT 666  
 Db 196 AsnAspPheTyrValThrGlySerTyrAlaGlyLysTyrValProAlaIleAlaHis 215  
 QY 667 CATTTTTCAGCAGAGATGAGAGAAATGCCAAGGTAGTGTAAATGGGTTCACGCTTAAI 726  
 Db 216 LeuIleHisSerLeuAsn-----ProValArgGluValLysIleAsn 229  
 QY 727 TTCACACTCTCGGGAATTAATACGGCATCATCGACGAGGAGATCCAGGCCCTTACAC 786  
 Db 230 LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyGlyTyr 249

QY 787 CCTGAATTCCTGTGAACAATACCTACGGTATCAAGGTGTCAACGAGACCGCTCTACAAC 846  
 Db 250 AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGlyLysGlnLysLys 265  
 QY 847 TACATGAAGTTTGCACCAACCAATGCAATGGTTGCAGGATTTGATTCACCTCAAA 906  
 Db 266 Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg 279  
 QY 907 CAGACAAAC-----CGCACCGCATTAGCTAGCTACGCCCTCTGCCCGCAGGACAC 957  
 Db 280 LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr 299  
 QY 958 AAC-----ATGTGACAGGACAATGTTGAGGG-----CCATACACTACGCCCTTTCGCTGGT 1005  
 Db 300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu--- 318  
 QY 1006 CGTGGTGTGTATGATTCGGCATCATATGATGACCCGACTCGGCCAACTTTATACAAC 1065  
 Db 319 -----ArgCysThrGluProGluAspGlnLeuTyrTyrVal 330  
 QY 1066 AAATTTCTGGCAAGGACTCTGTCAAGCGCTATCGCGCTCAACATCAACTACACCCAG 1125  
 Db 331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349  
 QY 1126 TCCAATAAATGACGCTTAC-----TAGCTTTTCCAGCAACAGCGCGACTTGTCTGGCC 1179  
 Db 350 AsnAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLysPro 369  
 QY 1180 AACTTCATCAAGACCTCGAGGAGATCTTTCCTCCCGTGGTGTCTCCCTCATCTAT 1239  
 Db 370 TrpLeuThrGluLeuMetAsn-----TyrLysValLeuIleTyrAsn 384  
 QY 1240 GCGGACGCGCATATCATC-----TGCAAC 1263  
 Db 385 GlyGlnLeuAspIleIleValAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp 404  
 QY 1264 TGGTTCGGGTGAGCGCGTTCCTCGTCCGAACTACTCCCAAGCGGCCAG----- 1317  
 Db 405 TriPysGlySerGln-----GluTyrLysLysAlaGluLysLysVal 418  
 QY 1318 -----TWCCGAAGC-----GCAGGTACAGCGCCCTGAAAGTCAACGGC 1356  
 Db 419 TriPysIlePheLysSerAspSerGluValAlaGlyTyr----- 431  
 QY 1357 GTCCAGTATGGGAACTCGGAGTATGTAATTTCTCTTCACTCCGCTCTATCAGGCA 1416  
 Db 432 -----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlyGly 446  
 QY 1417 GGCCATGAAGTCCCATACTACCGCCATCCCGCCCTCCCTGCAATTTGTTTAAACCGGACTATC 1476  
 Db 447 GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle 466  
 QY 1477 TTC-----GGTGGGAT 1488  
 Db 467 TyrGlyLysGlyTyrPasp 472

## RESULT 12

US-09-906-838-164  
 ; Sequence 164, Application US/09906838  
 ; Publication No. US20030027143A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Genentech, Inc.  
 ; APPLICANT: Ashkenazi, Avi  
 ; APPLICANT: Botstein, David  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Eaton, Dan L.  
 ; APPLICANT: Ferrara, Napoleone  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Fong, Sherman  
 ; APPLICANT: Gao, Wei-Giang  
 ; APPLICANT: Gerber, Hanspeter  
 ; APPLICANT: Gerritsen, Mary E.

```

; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tomas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/906,838
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-906-838-164

```

```

Alignment Scores:
Pred. No.: 1,29e-21 Length: 476
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.36% Indels: 116
DB: 11 Gaps: 21

```

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-09-906-838-164 (1-476)

QY 139 CCGSGT-----GTCAAAATCTACTCTGGATAT-----GTGACACCTCTCCGAGTCC 186

```

Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyLeuThrValAlaSerLysThrTyrAsnSer 86
QY 187 CATACCTCTCTTGGTCTTCTGGAAGCCAGACATAACCCAGAACTGACACCTATCACATTG 246
Db 87 AsnLeuPhePheIrrPhePheProAlaGlnIleGlnProGluAspAlaProValValLeu 106
QY 247 TGGTIGATGGTGGCCCTGGAGCGATCTTTTGAATCGGTCTCTTCGAGAGITGGCCCT 306
Db 107 TrpLeuGlnGlyGlyProGlyGlySerSerMetPheGlyLeuPheValGluHisGlyPro 126
QY 307 TGGCATGTCAATTCGACTTTTGTATCATACATCAACCCCTCACTCTCGAGAGGAGTCTCC 366
Db 127 TyrValValThrSerAsnMetThrLeuArgAspArgAspPhePhePheIrrThrLeu 146
QY 367 AATTACTATTCCTGTCCTCCAGCCATTCGGAGTCGGCTTTTCATATAGTATACGGTTGAT 426
Db 147 SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspAspThrHis 166
QY 427 GGGTCCATTAACCCCTGTAACCTGGGTCTGTCGAAATTCGAGCTTTCGAGAGTTCAGGGC 486
Db 167 Gly----- 167
QY 487 CGGTACCCCAACCATTTGAIGCCACITCTGATCGATACCTACCAATCTTGCCGAGAGGCCGCT 546
Db 168 -----TyrAlaValAlaSerLysThrValAlaArgAspLeuTyr 180
QY 547 IGGGAGATCCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTTGGAGTTCAGGTCT 606
Db 181 SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn 195
QY 607 AAGGACTTCAGTCTATGGACGGAGAGCTATGGAGGCGACATATGGTCTCGATTCCTCAAT 666
Db 196 AsnAspPheTyrValThrGlyGluSerTyrAlaGlyLysTyrValProAlaIleAlaHis 215
QY 667 CATTTTTCAGCAGAGATGAGAGAATTCGCAACGGTAGTGTGTTAATGGTGTGACGTTAAT 726
Db 216 LeuIleHisSerLeuAsn-----ProValArgGluValLysLeuAsn 229
QY 727 TTCAACTCTCTGGGAATTTATTAACGGCATCATCGACGAGCGGATCCAGGCCCTTACTAC 786
Db 230 LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyGlyTyr 249
QY 787 CCGTAATTCGCTGTGAACAATACCTACGGTAICAAAGGCTGTCAACGACCGCTCTACAAC 846
Db 250 AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGluLysGlnLysLys 265
QY 847 TACATGAAGTTTGCACCAACCAATGCCAAATGGTTCGACGAGATTTGATTTCCACCTGCAA 906
Db 266 Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg 279
QY 907 CAGACAAAC-----CGCACCCGCAATAGTACTACGCTCTGCGCCGCAAGCCAGC 957
Db 280 LysGlnAsnTrpPheGluAlaPheGluLeuAspLysLeuLeuAspGlyAspLeuThr 299
QY 958 AAC-----ATGTCGAGGCGACAATGTTGAGGGG-----CCATACTACCCCTTGTGCTGT 1005
Db 300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu--- 318
QY 1006 CGTGGTGTGTATGATATTCGGCAICCATATGATGATGACCGGACATCCCGCAAGTTATTACAAC 1065
Db 319 -----ArgCysThrGluProGluAspGlnLeuTyrVal 330
QY 1066 AAATTTCTGGCAAGACTCTGTCTATGGACGCTATCGCGCTCAACATCAACATACACCCAG 1125
Db 331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349
QY 1126 TCCATATATGAGCTCTAC-----TACGCTTTCAGCAACAGCGAGCTTGTGCTGGCC 1179
Db 350 AsnAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLysPro 369
QY 1180 AACITCATGAGACCTCGAGGAGATCCTGCTCTCCCGCTGCTGCTCCCTCATCTAT 1239

```

```

Db      370 TrpLeuThrGluIleMetAsnAsu-----TyrLysValLeuIleTyrAsn 384
QY      1240 GCGAGCGCGGATTACATC-----TGCAC 1263
Db      385 GlyGlnLeuAspIleValAlaAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp 404
QY      1264 TGGTTTCGGGGTCAGCGGTTTCCTCGGTGGAACTACTCCCAAGCCGCCAG----- 1317
Db      405 TrpLysGlySerGln-----GluTyrLysLysAlaGluLysLysVal 418
QY      1318 -----TTCGGAAGC-----GCAGGTACACGCCCTGAAAGTCAACGGC 1356
Db      419 TrpLysIlePheLysSerAspSerGluValAlaGlyTyr----- 431
QY      1357 GTCGAGTAGGGGAACATCCGAGTATGGTAATTTCTCTTCACTCGCGCTCATGGGCA 1416
Db      432 -----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlyGly 446
QY      1417 GGCATGAAGTCCCATACATACACAGCCATCGCTCCCTGCAATTTGTTAACCGGACTATC 1476
Db      447 GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle 466
QY      1477 TTC-----GGTGGGAT 1488
Db      467 TyrGlyLysGlyTyrAsp 472

RESULT 13
US-09-712-338-1_copy_55_1662 (1-1608) x US-09-712-338-1_copy_55_1662 (1-476)
; Sequence 164, Application US/0907613
; Publication No. US20030027145A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,613
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13

```

```

; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-712-338-1_copy_55_1662 (1-1608) x US-09-712-338-1_copy_55_1662 (1-476)

Alignment Scores:
Pred. No.: 1,29e-21 Length: 476
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.36% Indels: 116
DB: 21

US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-712-338-1_COPY_55_1662 (1-476)
QY 139 CCGGGT-----GTCAATCTACTCTGGATAT-----GTCCAGACCTCTCCGAGTCC 186
Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86
QY 187 CATACCTTCTCTGGTTCTTCGAGCCAGACATACCCAGAACTGCACCTATCACATTG 246
Db 87 AsnLeuPhePheTrpPheProAlaGlnIleGlnProGluAspAlaProValValLeu 106
QY 247 TGGTTGAATGGTGGCCCTGGAAGGATCTTTGATCGTCTCTTCGAGAGATTGGGCCT 306
Db 107 TrpLeuGlnGlyGlyProGlyGlySerSerMetPheGlyLeuPheValGluHisGlyPro 126
QY 307 TGCCATGTCAATTCGACTTTTGTGATCATCACTCAACCTCTACTCGTGAACGAGGTCTCC 366
Db 127 TyrValValThrSerAsnMetThrLeuArgAspArgAspPheProTyrThrThrLeu 146
QY 367 AATTACTATTCTCTGTCGCCAGCCATTCGGAGTCGGCTTTCATATAGTATACGTTGAT 426
Db 147 SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspAspThrHis 166
QY 427 GGGTCCATTAAACCTGTAACTGGGGTCTGTCGCAAAATTCGAGCTTTCGAGGAGTTCAGGC 486
Db 167 Gly----- 167
QY 487 CGGTACCCACACCATTTGATGCCACTCTGATCGATATACCAATCTTGGCCAGAGCCGCT 546
Db 168 -----TyrAlaValAsnGluAspAspValAlaArgAspLeuTyr 180
QY 547 TGGGAGATCTTCGAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGTCAGTCT 606
Db 181 SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn 195
QY 607 AAGGACTTCAGTCTATGACGAGGAGCTATGAGGACCTATGGTCTCTGCATCTCTCAAT 666

```

```

Db 196 AsnAspPheTyrValThrGlyGluSerTyrAlaGlyLysTyrValProAlaIleAlaHis 215
QY 667 CAFTTTTACGACGACAGATGAGAGAAATGCCAACGGTAGTGTTAATGCTTACCTTAAT 726
Db 216 LeuIleHisSerLeuAsn-----ProValArgGluValIleAsn 229
QY 727 TTCACACTCTGGGAATTTAATACGGCATCATCGCAGGCGATCCAGGCCCTTACTAC 786
Db 230 LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyGlyTyr 249
QY 787 CCTGAATTCGCTGGAACAATACCTACGGTATCAAGCTGTCAACGAGACGGCTTACAAAC 846
Db 250 AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGluLysGlnLysLys 265
QY 847 TACATGAAGTTTGCACCAAAATGCCAAATGGTTCGCCAGGATTTGATTTCCACTGCAAA 906
Db 266 Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg 279
QY 907 CAGACAAC-----CCACCGCATTTAGCTGACTACGCCCTCTCGCGCAGGCCACC 957
Db 280 LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr 299
QY 958 AAC-----AATGTCAGGGAATGTGAGGG-----CCATACACTACGCTTGTGCTGGT 1005
Db 300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu-- 318
QY 1006 CGTGGTGTGTATGATATTCGGCATCATATGATGACCGCAGCTCCGCCAAGTTTATACAA 1065
Db 319 -----ArgCysThrGluProGluAspGlnLeuTyrTyrVal 330
QY 1066 AATTTCTGCAAGGACTCTGTCAATGGAGCTATCGCGCTCAACATCAACTACACCCAG 1125
Db 331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349
QY 1126 TCCTAATATGAGCTTAC-----TAGCGTTTCCACGACACAGCGGACCTTGTCTGGGCC 1179
Db 350 AsnAspGlyThrIleValGlnLysTyrLeuArgGluAspThrValGlnSerValLysPro 369
QY 1180 ACTTTCATCGAAGACCTCGAGGAGATCCTTGTCTCCCGTGTGCTGCCCTCATCTAT 1239
Db 370 TrpLeuThrGluIleMetAsn-----TyrLysValLeuIleTyrAsn 384
QY 1240 GCGCAGCGCATATCATC-----TGCACAC 1263
Db 385 GlyGlnLeuAspIleIleValAlaLaLaLeuThrGluArgSerLeuMetGlyMetAsp 404
QY 1264 TGTTCGCGCGTCCAGCGCGTTTCCCTCGCTGCGAATCTCCCAAGCCGCCAG----- 1317
Db 405 TrpLysGlySerGln-----GluTyrLysLysAlaGluLysLysVal 418
QY 1318 -----TTCGGAAGC-----GCAGGTACACGCCCTCGAAGTCAACGCGC 1356
Db 419 TrpLysIlePheLysSerAspSerGluValAlaGlyTyr----- 431
QY 1357 GTCGAGTATGGGAACACCGCAGATGGTAATTTCTCCTTCACTCGCGCTCATGAGGCA 1416
Db 432 -----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlyGly 446
QY 1417 GCCCATGAAGTCCCATCTACCGCCCATCGCTCCCTGCAATTTGTTAACCGGACTATC 1476
Db 447 GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle 466
QY 1477 TTC-----GGTGGGAT 1488
Db 467 TyrGlyLysGlyTrpAsp 472

```

## RESULT 14

```

US-09-907-942-164
; Sequence 164, Application US/0907942
; Publication No. US20030027146A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi

```

```

Alignment Scores:
Pred. No.: 1.29e-21 Length: 476
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 52

```

```

; APPLICANT: Boistein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Pacini, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,942
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-942-164

```





Search completed: September 16, 2003, 18:56:17  
Job time : 211 secs

**This Page Blank (uspto)**

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	894	30.8	423	2	S38953	carboxypeptidase D	
2	746.5	25.7	460	1	S51516	serine-type carbox	
3	652.5	22.5	1002	2	T43236	carboxypeptidase C	
4	636	21.9	508	1	S45008	probable carboxype	
5	601	20.7	552	2	JC7666	serine-type carbox	
6	596	20.5	523	1	JC1380	carboxypeptidase C	
7	594.5	20.5	523	1	S61713	carboxypeptidase C	
8	527.5	18.2	532	1	CPBYV	carboxypeptidase C	
9	473	16.3	429	1	T03607	probable carboxype	
10	452	15.6	500	1	S22530	carboxypeptidase C	
11	439	15.1	510	2	I48977	carboxypeptidase-1	
12	425.5	14.7	411	1	A35275	carboxypeptidase C	
13	413.5	14.2	499	1	A29412	carboxypeptidase C	
14	413.5	14.2	2105	2	T18968	probable serine-ty	

```
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	894	30.8	423	2	S38953	carboxypeptidase D	
2	746.5	25.7	460	1	S51516	serine-type carbox	
3	652.5	22.5	1002	2	T43236	carboxypeptidase C	
4	636	21.9	508	1	S45008	probable carboxype	
5	601	20.7	552	2	JC7666	serine-type carbox	
6	596	20.5	523	1	JC1380	carboxypeptidase C	
7	594.5	20.5	523	1	S61713	carboxypeptidase C	
8	527.5	18.2	532	1	CPBYV	carboxypeptidase C	
9	473	16.3	429	1	T03607	probable carboxype	
10	452	15.6	500	1	S22530	carboxypeptidase C	
11	439	15.1	510	2	I48977	carboxypeptidase-1	
12	425.5	14.7	411	1	A35275	carboxypeptidase C	
13	413.5	14.2	499	1	A29412	carboxypeptidase C	
14	413.5	14.2	2105	2	T18968	probable serine-ty	

[illegible]

15	383	13.2	482	2	T49079	serine-type carboxypeptidase C
16	363	12.5	480	2	T50311	serine-type carboxypeptidase C
17	362.5	12.5	487	2	T49080	hypothetical protein
18	358	12.3	470	2	B96637	hypothetical protein
19	352.5	12.1	470	2	T16606	probable serine carboxypeptidase
20	351	12.1	482	2	A43828	probable serine carboxypeptidase
21	349	12.0	474	2	A35732	probable serine carboxypeptidase
22	348	12.0	452	2	H84772	protective protein
23	344.5	11.9	516	2	S44191	probable serine carboxypeptidase D
24	344	11.8	465	2	G86244	serine carboxypeptidase D
25	344	11.8	501	2	T49081	serine carboxypeptidase D
26	340	11.7	512	2	T33463	serine-type carboxypeptidase
27	337.5	11.6	425	2	F85360	probable serine carboxypeptidase
28	333.5	11.5	510	2	T39601	SERINE CARBOXYPEPTIDASE
29	333	11.5	487	2	B84472	serine carboxypeptidase
30	332.5	11.4	480	2	A33589	probable serine carboxypeptidase C
31	332	11.4	465	2	B85358	SERINE CARBOXYPEPTIDASE C
32	332	11.4	469	2	T16316	SERINE CARBOXYPEPTIDASE C
33	330.5	11.4	474	2	D84631	probable serine carboxypeptidase
34	330	11.4	1203	2	T21275	probable serine carboxypeptidase
35	327.5	11.3	479	2	G96501	hypothetical protein
36	327.5	11.3	2338	2	T25810	probable serine carboxypeptidase
37	326.5	11.2	425	2	E84631	hypothetical protein
38	326.5	11.2	454	2	T20829	probable serine carboxypeptidase
39	321.5	11.1	729	2	A23651	probable serine carboxypeptidase
40	320.5	11.0	456	2	H86406	KEX1 protein precursor
41	319	11.0	502	2	T49188	probable serine carboxypeptidase
42	318	11.0	468	2	D84503	serine carboxypeptidase
43	316	10.9	574	2	T16230	probable serine carboxypeptidase
44	314.5	10.8	476	1	T50701	hypothetical protein
45	312	10.7	510	1	S43516	carboxypeptidase D

Db 46 AlaAlaIrrPpHeAsnGlyProGlyCysSerSerMetIleGlyLeuPheGlnGluAsn 65  
 QY 301 GGCCCTTCCCAT---GTCAATTCGACTTTTGTATGACTACATCAACCCCTCACTCGTGGAC 357  
 Db 66 GlyProCysHisPheValAsnGlyAspSerThrProSerLeuAsnGluAsnSerIrrPpAsn 85  
 QY 358 GAGGTCTCCATTTACTATTCTCTCCAGCCATTTGGAGTGGCTGCTTTTCATATAGTAT 417  
 Db 86 AsnTyrAlaAsnMetIleTyrIleAspGlnProIleGlyValGlyPheSerTyrGly--- 104  
 QY 418 ACGGTTGATGGTCCATTAAACCCCTGTAACCTGGGGTGTGCGAAATTCGAGCTTTCAGGA 477  
 Db 104 ----- 104  
 QY 478 GTTCAGGGCGGTACCCACCAATTCATCCACTCTGATCGATACCTACCAATCTGCCGA 537  
 Db 105 -----ThrAspAspValThr-----SerThrValThrAlaAla 115  
 QY 538 GAGGCGGCTTGGGAGATCTCTCAAGGATTCCTTAGTGGACTACCTAGCTGGACTCTAGG 597  
 Db 116 ProTyrValTrpAsnLeuLeuGlnAlaPheTyrAlaGlnArgProGlnTyrGluSerAsp 135  
 QY 598 GTCCAGCTTAAGAGCTTCACTATGTCGACGGAGAGCTATGAGGGGACATAGTCTCGCA 657  
 Db 136 -----AspPheAlaIlePheThrGluSerTyrGlyGlyHisTyrGlyProGlu 151  
 QY 658 TTCTTCAATCATTTTACGAGCAGATGAGAGAAATGCCAAGTAGTGTATATGCTGT 717  
 Db 152 PheAlaSerTyrIleGlnGlnAsnAlaAlaIleLysAlaGlySerValThrGlyGln 171  
 QY 718 CAGCTTAATTAACCTCTGGGAATTTATTAAGGCATCATCGACGAGCGATCCAGGCC 777  
 Db 172 AsnValAsnIleValAlaLeuGlyValAsnAsnGlyTrpIleAspSerThrIleGlnGlu 191  
 QY 778 CCTTACTACCTGAATTCGCTGGAACAATCACTACGCTATCAAGGCTGTCAAGAGAAC 837  
 Db 192 LysAlaTyrIleAspPheSerTyrAsnAsnSerTyrGlnIleIleAspSerSerThr 211  
 QY 838 GTCTCAACTACATGAAGTTGTCACCAAAATGCCAAATGGTGGCCAGGATTTGATTCC 897  
 Db 212 ArgAspSerLeuLeuAspAlaTyrAsn-----AsnGlnCysLeuProAlaLeuGln 228  
 QY 898 ACCTGTCAAGAACAAACCCAGCCAGCTAGCTACGCTCTCCGCGGAGCCAGCC 957  
 Db 229 GlnCysSerGlnSerGlySerThr-----SerAspCysThrAsnAlaAsp 243  
 QY 958 ACATGTGAGGAGCAATGTGAGGGCCATACCTACCCCTTGTGCTGCTGTGTAT 1017  
 Db 244 SerValCysTyrGlnAsnIleGluGlyProIleSerSerGlyAspPheAspValTyr 263  
 QY 1018 GATATTGGCATCCATATGATGACCCGACCTCCGCAAGTTATACAAATTTCTGGCA 1077  
 Db 264 AspIleArgGluProSerAsnAspProTyrProIleSerSerGlyAspPheLeuSer 283  
 QY 1078 AAGCACTCTGATGAGCGTATCGCGCTCAACATCAACATACCCGATCAATATGAC 1137  
 Db 284 AspProThrValValIysAlaIleGlyAlaArgThrAsnTyrGlnGluCysProAsnGly 303  
 QY 1138 GTCTACTACCTTTCAGCAAAACAGGCGACTTTCTCGGCCCACTTCATCGAGAGCTC 1197  
 Db 304 ProTyrAsnLysPheAlaSerThrGlyAspAsnPro---ArgSerPheLeuSerThrLeu 322  
 QY 1198 GAGGAGATCTGCTCTCCCGTGTCTCCCTCATCTATGCGGACCGCGATTAATC 1257  
 Db 323 SerSerValValGlnSerGlyIleAsnValLeuValTrpAlaGlyAspAlaAspTrpIle 342  
 QY 1258 TGCAACTGTTTCGCGGTTCAGGCGTTTCCCTGCTCGCACTACTCCCAAGCCGCCAG 1317  
 Db 343 CysAsnTrpLeuGlyAsnTyrGluValAlaAsnAlaValAspPheProGlyAsnAlaGln 362  
 QY 1318 TTCGAGACGCGGATACACGCCCTCGAAGTCAACGGCTGAGTAGTAGGGAACCTGCC 1377  
 Db 363 PheSerAlaLeuAspLeuAlaProTyrThrValAsnGlyValGluLysGlyGlnPheLys 382

QY 1378 GAGTATGGTAATTTCTCTCTACTCCGCTCTGTATGAGGAGCCATCAAGTCCCATCTAC 1437  
 Db 383 ThrValAspAsnPheSerPheLeuLysValTyrGlyAlaGlyHisGluValProTyrTyr 402  
 QY 1438 CAGCCCATCGCTCCCTCGAATTTTAAACCGGACTATCTCGGTGGGATATCGCAGAG 1497  
 Db 403 GlnProAspThrAlaLeuGlnAlaPheLysGlnIleIle----- 415  
 QY 1498 GCCCAGCAAGAG 1509  
 Db 416 ---GlnLysLys 418  
 RESULT 2  
 S51516  
 serine-type carboxypeptidase (EC 3.4.16.-) Z precursor - Absidia zychaee  
 C:Species: Absidia zychaee  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 16-Jun-2000  
 C:Accession: S51516; S78013; S78014  
 R:Lee, B.R.; Takeuchi, M.; Kobayashi, Y.  
 Curr. Genet. 27, 159-165, 1995  
 A:Title: Molecular cloning and sequence analysis of the scpZ gene encoding the serine  
 A:Reference number: S51516; MUID:95308538; PMID:7788719  
 A:Accession: S51516  
 A:Molecule type: DNA  
 A:Residues: 1-460 <LEE>  
 A:Cross-references: EMBL:D16519; NID:g556466; PIDN:BA03966.1; PID:g995456  
 A:Experimental source: strain NRIC 1199  
 A:Accession: S78013  
 A:Molecule type: protein  
 A:Residues: 52-62; 90-99; 367-381 <LEE>  
 A:Accession: S78014  
 A:Molecule type: mRNA  
 A:Residues: 18-460 <LES>  
 C:Genetics:  
 A:Gene: scpZ  
 A:Introns: 64/3; 95/1; 107/2; 138/3; 234/2; 246/1; 270/1; 343/3; 353/2; 384/3; 437/3  
 C:Superfamily: serine carboxypeptidase  
 C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase  
 F:1-17/Domain: signal sequence #status predicted <SIG>  
 F:18-51/Domain: propeptide #status predicted <PRE>  
 F:52-460/Product: serine-type carboxypeptidase #status experimental <MAT>  
 F:40,46,119/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:186/Active site: Ser #status predicted  
 Alignment Scores:  
 Pred. No.: 9.44e-43 Length: 460  
 Score: 746.50 Matches: 175  
 Percent Similarity: 50.31% Conservative: 71  
 Best Local Similarity: 35.79% Mismatches: 182  
 Query Match: 25.71% Indels: 61  
 DB: 1 Gaps: 14  
 US-09-712-338-1\_COPY\_55\_1662 (1-1608) x S51516 (1-460)  
 QY 16 CCGCGCTCCGCTCGTAGAGACAGTACCCCAAGAACCCCGGGGTCAAGACTCTTACA 75  
 Db 29 ProMetSerHisValLeuArgGlnAspGlyAsnAspThr-----Ser 43  
 QY 76 ACCGCAACAATGTCACCATCCGCTACAGGAACCCGCGGAGAGGGCGTCTGGAGACT 135  
 Db 44 SerGlyAsnThrThrGlnLeuLysTyrThrSerPro-----LysLeuCysAsp--- 59  
 QY 136 ACCCGGGTGTCAATCTCTACTCTGTGATATGTGACACTCTCGGAGTCCCATCTTC 195  
 Db 60 ---ProAspValLysGlnTyrSerGlyTyrLeuAspAlaAlaAsnAspGluHisTyrPhe 78  
 QY 196 TTCGGTCTTCGAGACCCAGACATACCCAGAACTGCACCTATCACATGTGTGAAT 255  
 Db 79 PheTyrPhePheGluSerLysAsnAspProLysAsnAspProLeuThrIleIrrPleuAsn 98  
 QY 256 GGTGGCCCTCGAAGCGATTCTTTGATCGGTCTCTTCGAGAGATTGGCCCTTGCCTATGTC 315

Db 99 GlyGlyProGlyCysSerSerLeuIleGlyLeuTrpGluGluLeuGlyProCysGlnGln 118  
QY 316 AATTCGACTTTTGATGACTACATCAACCCCTCAGCTCGTGGACGAGGCTCCAAATTACTA 375  
Db 119 AsnGlySer-----AlaAsnProHisSerIrrpHisSerSerAsnMetLeu 134  
QY 376 TTCCTGTCCAGCAGCATTTGGAGTGGCTTTTCATATATAGTGATACGGTTGATGGGTCCAIT 435  
Db 135 PhePheAspGlnProAspGlyValGlyPheSerIrr----- 146  
QY 436 AACCCCTGTAACCTGGGTCGTCGAAATTCAGAGCTTTCAGAGATTACAGGCGCGGTACCCA 495  
Db 147 -----GlyIysGln----- 149  
QY 496 ACCAATTGATGCCACTCTGATGATACACAAATCTCCCGCAGAGCGCGCTGGAGATC 555  
Db 150 -----ThrValSerThrThrGluAspAlaGluArgAlaIrrpPhe 164  
QY 556 CTCGAGGATTCCTTAGTGGACTACCTAGCTTAGCTAGCTAGCTAGCTAGCTAGCTAGCT 615  
Db 165 LeuGlnAlaPheThrGluThrPhePro-----GlnIrrSerIrrLeuAspVal 180  
QY 616 AGCTATGTGACGACGAGCTATGAGGCGCACTATGGCTGCATCTTCATCAATCTTTIAC 675  
Db 181 HisTyrrPheGlyGluSerIrrGlyIrrHisTyrrIleProGlyPheAlaSerHisValVal 200  
QY 676 GAGCAGAATGAGAAATTCGCAACGGTAGTGTAAATGGTCTTCAGCTTAATTTCAACTCT 735  
Db 201 AspMetAsnIrrIrrValGlnSerGlyGluGlyValValProLeuIrrSer 220  
QY 736 CTGGGAATATTAAACGGCATCATCGAGGCGGATCCAGCCCTTACTACCTGATTC 795  
Db 221 IleGlyValGlyAsnGlyPheIleAspAlaValIleGlnIrrIrrSerIrrProIrrMet 240  
QY 796 GCTGTGCAACAATACCTACCGATCAAGCTGTCAAGCAGACCTCTACACTACATGAAG 855  
Db 241 ThrCysHisSerIrrIrrIrrProAlaValLeuSerGluGlu-----GluCysAspLys 257  
QY 856 TTTGCCAACCAATGCTGCGAGATTTCAGGATTTCACCTGCAACAGACAAAC 915  
Db 258 MetGlnGlnIrrIrrGluAsnAspCysIrrIrrProAlaAlaGluGlnCysAlaGluSerAsp 277  
QY 916 CGCACCGCATTAGCTGACATACGCGCTCGCGCGAAGCCACCAACATGTGCGGGACAAT 975  
Db 278 -----GluAspSerAspCysValAsnAlaAsnGlnCys---GlyGln 291  
QY 976 GTTAGGGCGCATACCTACCGCTTTCGTGCTGCTGTGTATGATATTCGGCAICCAT 1035  
Db 292 IleGluGly---IleIrrAlaGlnSerGlyIrrSerPheIrrAspIrrArgGlnGlnGly 310  
QY 1036 GATGACCGGACGCTTCGTCGCGCCCACTTACAAATTTCTGCGAAAGGACTGTGTCATGGAC 1095  
Db 311 AspAsp---ThrProHisProPheValAspGlu---LeuAsnIrrIrrAlaSerValIrrIrrLys 328  
QY 1096 GCTATCGCGGTCAACATCACTACCCAGTCCATATATAGCTCTACTACGTTTCCAG 1155  
Db 329 GluValGlyAlaArgGlyHisPheSerMetCysSerAspSerValGlyThrAlaPheAla 348  
QY 1156 CAACAGCGGACTTTGCTGCGCCCACTTCATCGRAGACCTCGAGAGATCTTGTCTCTC 1215  
Db 349 GlnThrGlyAspGlyAla---ArgSerIrrIrrProAlaValGluIrrIrrLeuIrrGlu 367  
QY 1216 CCGGTGCGGTCTCCCTCATCTATGCGGACGCGGATTACATCTGCACTGGTTCGGCGT 1275  
Db 368 GlyIleProValLeuIrrIrrValGlyAspAlaAspValIrrIrrIrrIrrIrrIrrIrr 387  
QY 1276 CAGCGCGTTCCTCGCTCGCAACTACTCCCAAGCGCCGACGTTCCGAAGCGCAGGGTAC 1335  
Db 388 LeuAspValAlaAspSerLeuIrrIrrIrrAspGlySerAlaPheSerIrrIrrIrrIrr 407  
QY 1336 ACGCCCTGAAAGTCAACGGCTCGAGTATGGGAACTCCGCGAGTATGGTAAATTTCTCC 1395  
Db 408 GluAlaTrpIrrAlaAspGlyIrrGluValGlnPheArgSerAlaAspIrrIrrIrr 427

QY 1396 TTCCTGCGCTCTATGAGGCGCATCAAGTCCCATACTACCCAGCCCATGCCCTCCG 1455  
Db 428 PheValArgValIrrGluAlaGlyHisGluValProMetTyrrGlnProGluAlaLeu 447  
QY 1456 CAATTGTTTTAACCGGACTATCTCGGI 1482  
Db 448 SerMetPheGlnThrTrpIleSerGly 456  
RESULT 3  
T43236  
carboxypeptidase C (EC 3.4.16.5) precursor [validated] - fission yeast (Schizosacchar  
N:Alternate names: carboxypeptidase Y  
C:Species: Schizosaccharomyces pombe  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 08-Sep-2000  
C:Accession: T43236; T37997  
R:Tabuchi, M.; Iwahara, O.; Ohtani, Y.; Ohuchi, N.; Sakurai, J.; Morita, T.; Iwahara  
J. Bacteriol. 179, 4179-4189, 1997  
A:Title: Vacuolar protein sorting in fission yeast: cloning, biosynthesis, transport,  
A:Reference number: Z2357; MUID:97352672; PMID:9209031  
A:Accession: T43236  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1002 <TAB>  
A:Cross-references: EMBL:D86560; NID:g3046860; PIDN:BAA25568.1; PID:g3046861  
R:Oliver, K.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
Submitted to the EMBL Data Library, June 1997  
A:Reference number: Z21760  
A:Accession: T37997  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1002 <OLI>  
A:Cross-references: EMBL:Z97209; PIDN:CA10121.1; GSPDB:GN00066; SPDB:SPAC19G12.10C  
A:Experimental source: strain 972h(-); cosmid c19G12  
C:Comment: The mature carboxypeptidase exists as a heterodimer held together by a dis  
C:Genetics:  
A:Gene: cpy1; SPAC19G12.10C  
A:Map position: 1  
C:Superfamily: serine carboxypeptidase  
C:Keywords: disulfide bond; glycoprotein; heterodimer; hydrolase; serine carboxypepti  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-521/Domain: propeptide #status predicted <PRO>  
F:200-332/Region: 13-residue repeats (H-H-[EK]-p-G-E-H-M-P-P-P-[MF])  
F:344-425/Region: 9-residue repeats (E-H-H-[RKO]-G-G-[KE]-[DE]-K)  
F:522-1002/Product: carboxypeptidase C #status experimental <MAT>  
F:627-880/Disulfide bonds: #status predicted  
F:659/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:715,921,978/Active site: Ser, Asp, His #status predicted  
Alignment Scores:  
Pred. No.: 2,62e-36 Length: 1002  
Score: 652.50 Matches: 165  
Percent Similarity: 47.34% Conservative: 66  
Best Local Similarity: 33.81% Mismatches: 174  
Query Match: 22.47% Indels: 83  
DB: 2 Gaps: 15  
US-09-712-338-1\_copy\_55\_1662 (1-1608) x T43236 (1-1002)  
QY 91 ACCATCCGTACAAAGAACCCGGGCGAGGGCGTCTCGGAGACTACCCGGGTGCAAA 150  
Db 570 ThrLeuArgValIrrAspSerIrrIrrProGluSerLeuIrrIrrIrrIrrIrrIrrIrr 587  
QY 151 TCCTACTCTGTGATGTGACACCTCTCCCGAGTCCCATACCTCTCTGTGTTCTTCGAA 210  
Db 588 GlnTyrrThrGlyIrrLeuAspValGluAspArgHisLeuPheIrrPheGlu 607  
QY 211 GCCAGACATAACCCAGAACTGCACCTATCATCTGTGTTGAATGGTGGCCCTGGAAGC 270  
Db 608 SerArgAsnAspProGluAsnAspProValValLeuIrrIrrIrrIrrIrrIrrIrr 627  
QY 271 GATTCTTTCATCGTCTCTTCGAGAGTGGCCCTTCGACATGTCAT---TCGACTTTT 327  
Db 271 GATTCTTTCATCGTCTCTTCGAGAGTGGCCCTTCGACATGTCAT---TCGACTTTT 327

Db 628 SerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIleAsnIleGluThrLeu 647  
 QY 328 GATGACTACATCAACCCCTCAGTCGCGGAGGAGTCTCCAAATTAATCTATCTGTCGCCAG 387  
 Db 648 LysProGluTyrAsnProHisSerIleThrPasnSerAsnAlaSerValIlePheLeuAspGln 667  
 QY 388 CCATTGGGAGTCGGCTTTTCATATAGTATGAGGGTTCATTAACCCCTGTAAC 447  
 Db 668 ProIleAsnThrGlyPheSerAsnGlyAsp-----AspSerValLeuAspThrValThr 685  
 QY 448 GGGGTGCTGAAAATTTCGAGCTTTCAGGAGTTCAGGCGGTACCCCAACCATTTGATGCC 507  
 Db 685 ----- 695  
 QY 508 ACTCTCATGATACATACCATCTTTCGCGAGAGCGCTTGGGAGATCCTGCAAGGATTC 567  
 Db 686 -----AlaGlyLysAspValTyrAlaPheLeuAsnLeuPhe 697  
 QY 568 CTTAGTGGACTACTAGCTTGGACTCTAGGGTCAGCTTAAGGAGCTTCACTCTATGGACG 627  
 Db 698 PheAlaLysPhePro-----GlnTyrAlaHisLeuAspPheHisIleAlaGly 713  
 QY 628 GAGAGCTAGGAGGGGACATATGCTGCAATCTTCAATCATTTTACGACGAGATGAS 687  
 Db 714 GluSerTyrAlaGlyHisTyrIleProGlnPheAlaLysGluIleMetGluHisAsnGln 733  
 QY 688 AGA-----ATTCGCAACGGT---AGTGTAAATGGTTCAGCTTAATTTCAAC 732  
 Db 734 GlyAlaAsnPhePheValAlaSerGlyTyrGluMetGluLysGlnTyrIleAsnLeuLys 753  
 QY 733 TCCTCGGAATTATTATTCAGTCATCAGCAGCGCATCCGCCCTTACTACCCCTGAA 792  
 Db 754 SerValLeuIleGlyAsnGlyLeuThrAspProLeuValGlnTyrPheTyrGlyLys 773  
 QY 793 TTCGCTGTGAACATACCTACGGT-----ATCAAG 822  
 Db 774 MetAlaCysGluSerProTyrGlyProIleMetSerGlnGluLysAspArgIleThr 793  
 QY 823 GCTGTCAACGAGACCGCTCAACAATCATCATCAAGTTTCCCAACCAATGCCAAATGGTTCG 882  
 Db 794 GlyAlaTyrAspThr-----Cys 799  
 QY 883 CAGGATTTGATTTCCACCTGCAACAGACACAAACCGCACCGCATAGCTAGTACGCCCTC 942  
 Db 800 AlaLysLeuIleThrGlyCysTyrGlnThrGlyPheThr-----ProVal 814  
 QY 943 TGCCCGCAACCCACCAACATCTGCGAGGACAAATGTCAGGGCCATACTACGCCCTTGTCT 1002  
 Db 815 CysIleGlyAlaSerLeuTyrCysAsnAlaMetIleGlyPro---PheIleLysThr 833  
 QY 1003 GGTGCTGTGTATGATATTCGGCATCCATATGATGACCGGACTCCGCCAAGTAT--- 1059  
 Db 834 GlyLeuAsnIleTyrAspIleArgGluLysArgAspGlnGluHisLeuLysCysTyrPro 853  
 QY 1060 -----TACACAAATTTCTGGCAAGGACCTCTGTCATGAGCGGTATCGGGCTC 1107  
 Db 854 GluThrGlyAlaIleGluSerTyrLeuAsnGlnGluPheValGlnGluAlaLeuGlyVal 873  
 QY 1108 AACATCACTACACCCAGTCCAAATAGTACGTCTACTACGCTTTCCACCAACAGGGGAC 1167  
 Db 874 GluTyrAspTyrLysGlyCysAsnThrGluValAsnIleGlyPheLeuPheLysGlyAsp 893  
 QY 1168 TTTCGTGGCCCACTTCATCGAAGACTCCGAGAGATCCCTP-----GCCTCCCGCTG 1221  
 Db 894 TrpMetArgLysThrPheArgAspAspValThrAlaIleLeuGluAlaGlyLeuProVal 913  
 QY 1222 CGTCTCCCTCATCTATP---GGCGACCGCATATCATCTCCCACTGTTCCGGCGGTACG 1278  
 Db 914 -----LeuIleTyrAlaGlyAspAlaAspTyrIleCysAsnTyrMetGlyAsnGlu 930  
 QY 1279 GCCGTTCCCTCGCTGCGAATCTATCCCAACCGCCGAGTTCGAGCGGAGGTACACG 1338  
 Db 931 AlaTrpThrAspAlaLeuGluTrpAlaGlyGlnArgGluPheTyrGluAlaGluLeuLys 950

## RESULT 4

S46008

probable carboxypeptidase C (EC 3.4.16.5) YBR139w - yeast (*Saccharomyces cerevisiae*)  
 N;Alternate names: hypothetical protein YBR1015; hypothetical protein YBR139w

C;Species: *Saccharomyces cerevisiae*

C;date: 26-Aug-1994 #sequence\_revision 05-Jan-1996 #text\_change 19-Apr-2002

C;Accession: S46008; S46581

R;Becam, A.M.; Herbert, C.J.; Nasr, F.; Slonimski, P.P.; Zagulski, M.

submitted to the Protein Sequence Database, August 1994

A;Reference number: S45995

A;Accession: S46008

A;Molecule type: DNA

A;Residues: 1-508 <BEC>

A;Cross-references: EMBL:Z36008; NID:g536435; PIDN:CAA85097.1; PID:g536436; GSPDB:GNO

A;Experimental source: strain S288C

R;Becam, A.M.; Cullin, C.; Grzybowska, E.; Lacroute, F.; Nasr, F.; Ozier-Kalogeropoul

Yeast 10(Suppl.A), SI-S11, 1994

A;Title: The sequence of 29.7kb from the right arm of chromosome II reveals 13 comple

A;Reference number: S46369; MUID:94378/17; PMID:8091856

A;Accession: S46581

A;Molecule type: DNA

A;Residues: 1-508 <BE2>

A;Cross-references: EMBL:X75891; NID:g496856; PIDN:CAA53497.1; PID:g496869

A;Experimental source: strain S288C

C;Genetics:

A;Cross-references: SGD:S0000343

A;Map position: 2R

A;Note: MIPS:YBR139w

C;Superfamily: serine carboxypeptidase

C;Keywords: hydrolase; serine carboxypeptidase; transmembrane protein

F;6-22/Domain: transmembrane #status predicted <TMS>

F;219,415,474/Active site: Ser, Asp, His #status predicted

## Alignment Scores:

Pred. No.:	2,95e-35	Length:	508
Score:	636.00	Matches:	167
Percent Similarity:	46.29%	Conservative:	70
Best Local Similarity:	32.62%	Mismatches:	197
Query Match:	21.90%	Indels:	78
DB:	1	Gaps:	14

US-09-712-338-1\_copy\_55\_1662 (1-1608) x S46008 (1-508)

QY 1 CTTCCAGGAAGTACACCGCGCTCGGTGCTGAGACAGACTA-----CCCAAG 48

Db 43 LeuProGlnAsnThrGlnGlnThrLeuLysLeuAspArgLeuAsnHisAspProLeu 62

QY 49 AACCCACCGGGTCAAGACTCTTACACCCCAACAAATGTCACCATCCGTACAGGAA 108

Db 63 PheThrThrPheIleSerSerValAspThrAspTyrSerLeuArgLeuArgThrValAsp 82

QY 109 CCGGGGCGAGGGCGTCTGCGAGACTACCCCGGTGTCAATCTACTCTGGATATCTC 168

Db 83 ProSerLysLeuGlyIle-----AspThrValLysGluTrpSerGlyTyrMet 98

QY 169 GACACCTCTCCGAGTCCCATACCTTCTTCGTCTTCGAGACCCAGACATACCCAGNA 228

Db 99 AspTyrLysAspSerLysHisPhePheTyrTrpPheGluSerArgAsnAspProAla 118

QY 229 ACTGCACCTATCACATTTGGTGTGAATGCTGSCCCTGGAGAGCGATTCTTTGATCGTCTC 288

Db	119	AsnAspProIleIleLeuThrLeuAsnGlyGlyProGlyCysSerSerPheThrGlyLeu	138
QY	289	TTCGAGAGATTGGGCGCTTCGCAGTGCACAAATCGCACTTTTGATGACTACATCAACCCCTCAC	348
Db	139	LeuPheGluLeuGlyProSerSerIleGlyAlaAspMetLysProIleHisAsnProTyr	158
QY	349	TCGTGGAACGGGTCTCCAAATTTACTATTTCCTCCAGCCATTTGGAGTCGGCTTTTCA	408
Db	159	SerTPAsnAsnAsnAlaSerMetIlePheLeuGluGlnProLeuGlyValGlyPheSer	178
QY	409	TATAGTGATACGGTGTGATGGTCCATTAAACCCCTGTAACCTGGGTGCTCGAAATTCGAGC	468
Db	179	TyrGlyAsp	181
QY	469	TTTGCAGGAGTTACGGCGCGTACCAACCACTTATGCCACTCTGAICGNACTACCAAT	528
Db	182	-----GluLysValSerSerThrLys	188
QY	529	CTTGCCGAGAGCGCGCTTGGGAGATCTCTGCAAGGATTCCTTAGTGGACTACTAGCTGTG	588
Db	189	LeuAlaGlyLysAspAlaTyrIlePheLeuGluLeuPhePheGluAlaPheProHisLeu	208
QY	589	GACTCTAGGGTCAGCTTAAGCACTTCAGCTATAGTGGACGGAGAGCATATGGAGGGCACTAT	648
Db	209	-----ArgSerAsnAspPheHisIleAlaGlyGluSerTyrAlaGlyHisTyr	224
QY	649	GGTCTGTGATCTTCAATCAATTTTACGAGCAGAAT--GAGAGATTGCCAACGGTAGT	705
Db	225	IleProGlnIleAlaHisGluIleValLysAsnProGluArg	239
QY	706	GTTAATGGTGTTCAGCTTAATTTCAACTCTCTGGGAATATTATACGGCATCATCGACGAG	765
Db	240	-----ThrPheAsnLeuThrSerValMetIleGlyAsnGlyIleThrAspPro	255
QY	766	CGCATTCAGGCCCTTACTACCTCCATTCGCTGTG---AACATACCTACGGTATCAAG	822
Db	256	LeulleGluAlaAspTyrTyrGluProMetAlaCysGlyLysGlyTyrHisProVal	275
QY	823	GCTGTCAACGAGACCGCTACAACTACATGAAGTTTGCAACCAAAATGCCAAATGGTTCG	882
Db	276	LeuSerSerGluGluCysGlyLysMetSerLysAlaAlaGlyArg	290
QY	883	CAGGATTTGATTTCCACTCGAACAGACAAACCGCACCGCATTTAGCTGACTACGCCCTC	942
Db	291	-----CysArgGluLeuAsnLysLeuCysTyrAlaSerLysSerSer	304
QY	943	-----TCCGCCGAAGCCACCACTATGCGAGGACAAATGTGAGGGGCCATACTAGCC	996
Db	305	LeuProCysIleValAlaThrAlaTyrCysAspSerAlaLeuLeuGluProTyrIleAsn	324
QY	997	TTTCTCGTGTGTGTATGATATTTCGGCATCCATATGATGACCG	1044
Db	325	---ThrGlyLeuAsnValIleAspIleArgGlyProCysGluAspAsnSerThrAspGly	343
QY	1045	-----ACTCGCCCAAGTTATTACAAATAATTTCTGGCAAGGACTCTGTCATGGAC	1095
Db	344	MetCysTyrThrGlyLeuArgTyrValAspGlnTyrMetAsnPheProGluValGlnGlu	363
QY	1096	GCTATCGGCTCAACATC---AACTACACCCAGTCCAAATTAATGACGTCIACTAGCTTTC	1152
Db	364	ThrLeuGlySerAspValHisAsnTyrSerGlyCysAspAsnAspValPheThrGlyPhe	383
QY	1153	CAGCAAAACGGCACTTTGTCTGGCCCACTTCATCGAAGACCTCGAGGAGATCCTTGCT	1212
Db	384	LeuPheThrGlyAspGlySerLysPro---PheGlnGlnTyrIleAlaGluLeuLeuAsn	402
QY	1213	CTCCCGTGGTGTCTCCCTCATCTATGGGACGGCATATACATCTGCAACTGGTTCGGC	1272
Db	403	HisAsnIleProValLeuIleTyrAlaGlyAspLysAspTyrIleCysAsnTyrLeuGly	422
QY	1273	GGTCAGGCCGTTTCCCTCGCTGCGAATCTACTCCCAAGCCCGCCAGTTCGGAACGCGAGG	1332

Db 423 AsnHisAlaTrpSerAsnGluLeuGluTrpIleAsnLysArgArgTyrGlnArgMet 442

QY 1333 TACAGGCC-----CTGAAGICAAAGCGCTGAGATGGGAAACTCCGAGTANGG 1396

Db 443 LeuArgProTrpValSerLysGluThrGlyGluLeuGlyGlnValLysAsnIyrGly 462

QY 1387 AATTTCCTCTACTCGGCTATGAGCAGGCCATGAAGTCCATCTACTACCGCCCATC 1446

Db 463 PropheThrPheLeuArgIleTyrAspAlaGlyHisMetValProTyrAspGlnProGlu 482

QY 1447 GCCTCCCTGCCTAATGCTTTAACCGGACTATCTCGGT 1482

Db 483 AlaSerLeuGluMetValAsnSerTrpIleSerGly 494

RESULT 5

JC7666

serine-type carboxypeptidase homolog precursor - *Emericella nidulans*

N:Alternate names: carboxypeptidase Y homolog

C:Species: *Emericella nidulans*, *Aspergillus nidulans*

C:date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001

C:Accession: JC7666

R:Ohsumi, K.; Matsuda, Y.; Nakajima, H.; Kitamoto, K.

A:Title: Cloning and characterization of the cpyA gene encoding intracellular carboxypeptidase

A:Reference number: JC7666; M01D:2133188; PMID:11440134

A:Accession: JC7666

A:Molecule type: mRNA

A:Residues: 1-552 <OHS>

A:Cross-references: DDBJ:AB051820

C:Comment: This protein is an intracellular carboxypeptidase, which is localized in v

C:Genetics:

A:Gene: cpyA

A:Introns: 165/1

F:1-15/Domain: signal sequence (or 1-18, or 1-19) #status predicted <SIG>

F:16-131/Domain: pro-sequence, vacuolar targeting signal (or 19-131, or 20-131) #stat

F:131-132/Region: cleavage site, by Arg-Ile #status predicted

Alignment Scores:

Pred. No.:	7,06e-33	Length:	552
Score:	601.00	Matches:	154
Percent Similarity:	47.34%	Conservative:	77
Best Local Similarity:	31.56%	Mismatches:	175
Query Match:	20.70%	Indels:	82
DB:	2	Gaps:	19

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x JC7666 (1-552)

QY 79 GCAACAATGTCAACCTCCGGTACAAGAACCGGGCAGAGGGCTGCGAGACTACC 138

Db 127 AlaTyrAspLeuArgIleLysThrAspSerSerLeuGlyIle-----Asp 143

QY 139 CGGGTGTCAAATCCTACTCTGATATGTCGACACCTCCOCC---GAGTCCCATACCTTC 195

Db 144 ProAspValLysGlnTyrThrGlyTyrLeuAspAspAsnGluAsnAspLysHisLeuphe 163

QY 196 TTCCTGGTCTTCGAAGCCACAGACATAACCCAGAACTCCACTATCACATTCGTGGTGAAT 255

Db 164 TyrTrpPheGluSerArgAsnAspProLysAsnAspProValValLeuTrpLeuAsn 183

QY 256 GTGGCCCTGGAGCGAATTCCTTTGATCGGCTCTTCGAAGAGTTGGCCCTTCGCATGTC 315

Db 184 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerile 203

QY 316 AATTCGACTTTTGATCACTACACCTCACTCGTGGACGAGGTCTCCAAATTTACTA 375

Db 204 AspGluAsnIleLysProValTyrAsnProTyrAlaTrpAsnSerAsnAlaSerValile 223

QY 376 TTCCTGTCGCCCATTTGGAGTCGGCTTTTCATATAGTCATACGGTGTGATGGTCCATT 435

Db 224 PheLeuAspGlnProValAsnValGlyTyrSerTyrSer----- 236

QY 436 AACCCIGTAACGGGTGCTCGAAATATCGAGCTTTCAGAGAGTTCAGGGCGGTACCCA 495





```

QY 448 GGGGTCGTGAAATTCAGAGTTTCAGGAGTTCAGGCGCGGTACCAACCAATTGATGCC 507
Db 231 -----Ser 231
QY 508 ACTCTGATCATCTACCAATCTTGGCGCAGAGCGCGCTTGGGAGATTCCTGCAAGGATTC 567
Db 232 GlnSerValSerAsnThrIleAlaAlaGlyLysAspValTyrAlaPheLeuGlnLeuPhe 251
QY 568 CTTAGTGGACTACTAGTCTAGGACTCTAGGCTCAGTCTAAGGACITTCAGTCTATGGAGC 627
Db 252 PheLysAsnPro-----GluTyrAlaAsnLeuAspPheHisIleAlaGly 267
QY 628 GAGAGCTATGAGGCGCACTATGCTCTGCAATCTTCAATCATTTTACGAGCAGAAATGAC 687
Db 268 GluSerTyrAlaGlyHisTyrIleProAlaPheAlaSerGluIleLeuThrHisProGlu 287
QY 688 AGAATTGCCAACCGTAGTGTATAGTGTTCACCTTAATTTCAACTCTCTGGGAATTAAT 747
Db 288 Arg-----AsnPheAsnLeuThrSerValLeu 296
QY 748 -----AAGGCGATCATCGACGAGCGATCCAGGCGCCCTTACACCTGAATTCGCTGIG 801
Db 297 IleGlyAsnGlyLeuThrAspProLeuValGlnTyrGluTyrTyrGluProMetAlaCys 316
QY 802 ACAATACCTTACGCTATCAAGGCTGTCAACGAGACGCTCTACAACTACATGAAGTTTGC 861
Db 317 GlyGluGlyGlyGluProSerValLeuGluProGluGluCysAspGlyMet-----Leu 334
QY 862 AACCAATGCCAATAGTTGCCAGGATTTGATTTCACCTGCAACAGACAAACCCGACCC 921
Db 335 AsnLeuLeuProArg---CysLeuSerLeuIleGluSerCysTyrGluSer----- 350
QY 922 GCATTAGCTGACTACGCCCTCTCGCGGAGCGACCAACATGTCGAGGGACAAATGTTGAG 981
Db 351 -----GlySerValTrpSerCysValProAlaThrIleTyrCysAsnAsnGlyGlnMet 368
QY 982 GGGCCATACGCGCTTCTGCTGGTGGTGTATGATATATTCGGCATCCATCATGATGAC 1041
Db 369 GlyProTyrGlnLys---ThrGlyArgAsnValTyrAspIleArgThrMetCysGluGly 387
QY 1042 CCG-----ACTCGGCCAGTTATACACAAATTTCTGGCAAGGACTCTGTC 1089
Db 388 SerSerLeuCysTyrSerGlnLeuGluTyrIleAspGlnTyrLeuAsnLeuProGluVal 407
QY 1090 ATGAGCGCTATCGCGGTCAACATCAAC---TACACCCAGTCCAAATACAGCTCTACTAC 1146
Db 408 LysLysAlaLeuGlyAlaGluValAspGluTyrGlnSerCysAsnPheAspIleAsnArg 427
QY 1147 GCTTTCCAGCAACAGCGGCACTTTCTCTGGCCC-----AACTTCATCGAAGACCTC 1197
Db 428 AsnPheMetPheAlaGlyAspTrpMetLysProTyrGlnLysAsnValIleAspLeuLeu 447
QY 1198 GAGGAGATCTTCTGCTCCCGCGGTGTCTCCCTCATCTAT---GGCGACCGCGATAC 1254
Db 448 GluLys-----GluLeuProVal-----LeuIleTyrAlaGlyAspLysAspPhe 462
QY 1255 ATCTGCAACTGTTGCGCGGTTCAGCGGTTTCCCTCGTCCGTAAGTCTCCCAAGCGCCG 1314
Db 463 IleCysAsnTrpLeuGlyAsnGlnAlaTrpThrAsnArgLeuGluTrpSerGlySerLys 482
QY 1315 CAGTTCCGAAGCGCGGTACACGCCCTCGAAGCTCAACGGCGCTCGAGTATGGGAAACT 1374
Db 483 GlyPheThrLysAlaProValLysSerTrpLysValGlyLysAsnAlaAlaGlyGluVal 502
QY 1375 CCGGAGTATGTAATTTCTCTACTCGCGTCTATGAGCGGCGCATGCAAGTCCCATAC 1434
Db 503 LysAsnTyrLysHisPheThrPheLeuArgValPheGlyGlyGlyHisMetValProTyr 522
QY 1435 TACAGCCCATCGCCTCCCTGCAATTTGTTAAACCGGACTATCTTCGGT 1482
Db 523 AspGlnProGluAsnAlaLeuAspMetValAsnArgTrpIleSerGly 538

```

## RESULT 7

```

S61713
carboxypeptidase C (EC 3.4.16.5) precursor - yeast (Pichia pastoris)
N:Alternate names: carboxypeptidase Y
C:Species: Pichia pastoris
C:Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 16-Jun-2000
C:Accession: S61713
R:Ohtani, H.; Ohtani, W.; Okazaki, N.; Furuhashi, N.; Ohmura, T.
Yeast 12, 31-40, 1996
A:Title: Cloning and characterization of the Pichia pastoris PRC1 gene encoding carbo
A:Reference number: S61713; MUID:96381245; PMID:8789258
A:Accession: S61713
A:Molecule type: DNA
A:Residues: 1-523 <OH>
A:Cross-references: EMBL:X87987; NID:g1171615; PIDN:CAA61240.1; PID:g1171616
C:Superfamily: serine carboxypeptidase
C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase; zymogen
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-106/Domain: propeptide #status predicted <PRO>
F:107-522/Product: carboxypeptidase C #status predicted <MAT>
F:193,271,484,487/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:249,441,498/Active site: Ser, Asp, His #status predicted
F:296-310,327-336/Disulfide bonds: #status predicted

Alignment Scores:
Pred. No.: 1,92e-32 Length: 523
Score: 594.50 Matches: 153
Percent Similarity: 45.89% Conservative: 59
Best Local Similarity: 33.12% Mismatches: 169
Query Match: 20.47% Indels: 81
DB: 1 Gaps: 16

US-09-712-338-1_COPY_55_1662 (1-1608) x S61713 (1-523)

QY 148 AAATCTACTCTGATATGTCGACACC---TCTCCCGAGTCCCATCACTTCTTCTGCTC 204
Db 121 LysGlnTyrSerGlyTyrLeuAspValGluAlaAspAspLysHisPheTyrTrpPhe 140
QY 205 TTCGAGCCAGACATACCCAGAACTGCACCTATCACATTTGGTGTGAATGGTGGCCCT 264
Db 141 PheGluSerArgAsnAspProGlnAsnAspProIleIleLeuTrpLeuAsnGlyGlyPro 160
QY 265 GGAAGCGATTTCTTGATCGTCTCTCGAAGACTTGGGCCCTTGCATGTCATTCGACT 324
Db 161 GlyCysSerSerLeuThrGlyLeuPhePheGluLeuGlySerSerArgIleAsnGluAsn 180
QY 325 TTGATGACTACATCAACCCCTCACTCGTGAAGAGGTCTCCCAATTTACTATTCCTGTC 384
Db 181 LeuLysProIlePheAsnProTyrSerTrpAsnGlyAsnAlaSerIleIleLysLeuAsp 200
QY 385 CAGCCATTTGGAGTCCGCTTTTCATATAGTATGATCGGTGTGATGGGTTCATTACCCCTGA 444
Db 201 GlnProValAsnValGlyPheSerTyrSer----- 210
QY 445 ACTGGGTCGTGCGAAATTCAGCTTTGTCAGGAGTTTCAGGCGGTACCCACCACTTGA 504
Db 210 ----- 210
QY 505 GCACTCTGATGATGATCACTACCAATCTTCCCGCAGAGCGCGCTTGGGAGATCTCGCAAGGA 564
Db 211 SerSerSerValSerAsnThrValValAlaGlyLysAspValTyrAlaPheLeuGlnLeu 230
QY 565 TTCCTTAGTGGACTACCTAGCTTGGACTCTAGGGTGGAGTCTAAGGACTTCACTAGTCTAG 624
Db 231 PhePheGlnHisPhePro-----GluTyrGlnThrAsnAspPheHisIleAla 246
QY 625 ACGGAGAGCTATGAGGCGCACTATGCTCTGCAATTCCTCAATCATTTTACAGCAGAAAT 684
Db 247 GlyGluSerTyrAlaGlyHisTyrIleProValPheAlaAspGluIleLeuSerGlnLys 266
QY 685 GAGAGAATTCGCAACGGTAGTGTATTAATGTTGCTCAGCTTAATTTCACTCTCTGGGAAT 744
Db 267 AsnArg-----AsnPheAsnLeuThrSerVal 275

```







```
QY 676 GAGCAGAAATGAGAGAAATGCGAAGCGTAGTGTATTATGGTGTGTACGCTTAATTTCAACTCT 735
Db :|||:|||||
231 GlnGlyAsnLys-----AlaAsn-----GluGlyIleHisIleAsnLeuLysGly 245
QY 736 CTGGGAATTTAATACGCATCATCGACGAGCGATCCAGGCCCTTACTACCTGAAATC 795
Db :|||:|||||:|||||
246 PheAlaIleGlyAsnGlyLeuThrAspProAlaIleGlnIleGlnIleGlnIleGlnIle 265
QY 796 GCTGTGACAACTACCTACGCTGTCACGAGCGTCTACACGAGCGTCTACACTACATGAG 855
Db :|||:|||||:|||||
266 AlaLeuAsp-----MetAsnLeuLysLysSerAspTyrAspArgIle--- 280
QY 856 TTGCCCAACCAATGCGCAGGATTTGATTTCCACCTGCAACACAGACAAAC 915
Db :|||:|||||:|||||
281 -----AsnLysPheIleProPheCysGluPheAlaIleLysLeuGlyThrAsnGly 298
QY 916 CGCACCAGCAATAGCIGACTACGCCCTCTGCGCGAAGCCACCAACATGTCAGGACAAT 975
Db :|||:|||||:|||||
299 Lys-----AlaSerCysMetAlaAlaTyrMetValCysAsnSerIle 312
QY 976 GTTGAGGGGCCAATACAGCCTTTGCTGCTGCTGTGTATGATATTCGGCATCCATAT 1035
Db :|||:|||||:|||||
313 PheSerSerIleMetLysLeuValGlyThrLysAsnTyrTyrAspValArgLysGluCys 332
QY 1036 GATGACCCGACTCGGCCAAGTTATTACAAC-----AAATTCTGCCAAAGGACTCTGTC 1089
Db :|||:|||||:|||||
333 GlnGlyLysLeuCysTyrAspPheSerAsnLeuGluLysPheGlyAspLysAlaVal 352
QY 1090 ATGGAGCTATCGGCTC---AACATCAACTACACCCAGTCCCAATATGACGCTACTAC 1146
Db :|||:|||||:|||||
353 LysGluAlaIleGlyValGlyAspLeuGluPheValSerCysSerThrThrValTyrGln 372
QY 1147 GCTTTCCGCAACACAGGCGACTTTGCTGCGCCCAACTTCATCGAA---GACCTCGAGGAG 1203
Db :|||:|||||:|||||
373 AlaMetLeuThrAsp-----TrpMetArgAsnLeuGluValGlyIleProAla 388
QY 1204 ATCTCTGCTCTCCCGTGGTGTCTCCCTCATCTATGCGGAGCGCGATTCATCTGCAAC 1263
Db :|||:|||||:|||||
389 LeuLeuGluAspGlyIleAsnValLeuIleTyrAlaGlyGlnTyrAspLeuIleCysAsn 408
QY 1264 TGGTTCGGCGGTGAGCGCCGTTTCCCTGCTGCGAAGTCAACGCGCTGCTCCCAAGCCCGCTCCGA 1323
Db :|||:|||||:|||||
409 TrpLeuGlyAsnSerArgTrpValHisSerMetGluTrpSerGlyGlnLysAspPheVal 428
QY 1324 AGCGAGGTATACGCGCCCTGAAAGTCAACGCGCTGAGTATGGGAAACTCGCGAGTAT 1383
Db :|||:|||||:|||||
429 SerSerHisGluSerProPheValValAspGlyAlaGluAlaGlyValLeuLysSerHis 448
QY 1384 GGTATTTCTCTCTACCTCGCTCTATGAGGAGCGCCATGAAAGTCCCATACCAAGCC 1443
Db :|||:|||||:|||||
449 GlyProLeuSerPheLeuLysValHisAsnAlaGlyHisMetValProMetAspGlnPro 468
QY 1444 ATCGCTCCCTGCAATTTGTTAAACCGG 1470
Db :|||:|||||:|||||
469 LysAlaSerLeuGluMetLeuArgArg 477
```

RESULT 11  
T48977  
carboxypeptidase-like protein F14D17.80 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 24-Oct-2000  
C;Accession: T48977  
R;Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.;  
submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z25008  
A;Accession: T48977  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-510 <JOK>  
A;Cross-references: EMBL:AL353992; GSPDB:GN00061; ATSP:F14D17.80  
A;Experimental source: cultivar Columbia; BAC clone F14D17  
C;Genetics:  
A;Gene: ATSP:F14D17.80

A;Map position: 3  
A;Introns: 115/2; 171/3; 211/3; 259/3; 310/1; 341/3; 428/1; 472/3  
C;Superfamily: serine carboxypeptidase  
F:231,421,478/Active site: Ser, Asp, His #status predicted  
Alignment Scores:  
Pred. No.: 6,52e-22 Length: 510  
Score: 439.00 Matches: 130  
Percent Similarity: 42.43% Conservative: 83  
Best Local Similarity: 25.90% Mismatches: 191  
Query Match: 15.12% Indels: 98  
DB: 2 Gaps: 16  
US-09-712-338-1\_COPY\_55\_1662 (1-1608) x T48977 (1-510)

```
QY 16 CCGCGCTCCGTCGGTAGAAGACAGCTACCCCAAGAACCACCCGGGTCAAGACTCTTACA 75
Db :|||:|||||:|||||
82 ProAlaAlaValAspArgArg----- 88
QY 76 ACCGCCAAACAATGTCACCATCCGGTACAGGAACCCGGGCGAGAGCGTCTCGAGACT 135
Db :|||:|||||:|||||
89 -----GlySerGlyGly----- 92
QY 136 ACCCGGGTGTCAAATCTCTAC-----TCTGGATAT-----GTCGACACTCTCC 180
Db :|||:|||||:|||||
93 SerProSerValGlnAspPheGlyHisAlaGlyTyrTyrLysLeuProAsnSerLys 112
QY 181 GAGTCCCATACCTCTCTTCIGTTCCTTCGAAAGCGACATACCCAGAACTGCACCTATC 240
Db :|||:|||||:|||||
113 AlaAlaArgMetPheTyrPhePheGluSerArgThrAsn---LysAlaAspProVal 131
QY 241 ACATTTGGTGAATGTGGCCCTGGAGCGATTTCTTGATCGTCTCTTCGAAAGAGTTG 300
Db :|||:|||||:|||||
132 ValIleTrpLeuThrGlyGlyProGlyCysSerSerLeuLeuAlaLeuPheTyrGluAsn 151
QY 301 GGCCCTTGCCATTCGACTTCGATTTGATGATCATCAACCTCTCTCGTGGACGAG 360
Db :|||:|||||:|||||
152 GlyProPheThrValSerAsnAsnSerSerLeuSerTrpAsnGluPheGlyTrpAspLys 171
QY 361 GTCTCCAAITTAATCTCTGTCAGCCATTCGGAGTTCGGCTTTCATATAGTGATACG 420
Db :|||:|||||:|||||
172 AlaSerAsnLeuIleTyrValAspGlnProValGlyThrGlyPheSerTyrThrSerAsp 191
QY 421 GTTGATGGTCCATTAACCTGTAACTGGGTGCTCGAAAATTCGAGCTTTGAGAGGT 480
Db :|||:|||||:|||||
192 GlnSerAspLeuArgHisAspGluAspGlyVal----- 202
QY 481 CAGGGCCGTACCAACCATTTGATGCCACTCTCATCGATACCAATCTGCCCGCAGAG 540
Db :|||:|||||:|||||
203 -----SerAsn 204
QY 541 GCGCTTGGGAGATCCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTTAGGGG 600
Db :|||:|||||:|||||
205 AspLeuTyrAspPheLeuGlnAlaPhePheLysGluHisPro-----GlnPhe 220
QY 601 CAGTCTAAGGACTTCAGTCTATGGAGCAGACTATGGAGGCACACTATGCTGCTCATC 660
Db :|||:|||||:|||||
221 ValLysAsnAspPheTyrIleThrGlyGluSerTyrAlaGlyHisTyrIleProAlaLeu 240
QY 661 TTTCAATCATTTTACGAGCAGAGTATGCCAGGTTGCAAGGTAGTGTATGCTGTTCAG 720
Db :|||:|||||:|||||
241 AlaSerArgValHisArgGlyAsnLys-----AsnLysGluGlyThrHis 255
QY 721 CTTAATTTCAACTCTCTGGGAATTTAATACGGCATCATCGAGCGGCGATCCAGCCCT 780
Db :|||:|||||:|||||
256 IleAsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAsnProGluIleGlnTyrGly 275
QY 781 TACTACCTGAAATTCGCTGTGAACAATACTACGCTATCAAGGTGTCAAGCGCTC 840
Db :|||:|||||:|||||
276 AlaTyrAlaAspTyrAlaLeuAspMetAsnLeuIleThrGlnSerAspHisAsnLeu 295
QY 841 TACAACATCATGAAGTTTGGCAACCAATGTTGCCAGGATTTGATTTCCACC 900
Db :|||:|||||:|||||
```



```

Db      287 AspileGluPheValSerCysSerThrSerValTyrGlnAlaMetLeuThrAsp----- 304
QY      1168 TTITGCTGGCCCAACTTCATCGAA--GACCTCGAGGAGATCTTCTCTCCCGGTGGT 1224
Db      305 -----TrpMetArgAsnLeuGluValGlyIleProAlaLeuLeuGluAspGlyIleAsn 322
QY      1225 GTCCTCCCTCACTATGCGAGCGCGGATTCATCATCTGCACTGGTTCGCGGGTCAGCCCGTT 1284
Db      323 ValLeuIleTyrAlaGlyGluTyrAspLeuIleCysAsnTrpLeuGlyAsnSerArgTrp 342
QY      1285 TCCTCTGCGGAACACTACTCCAGCCCGCCAGTTCGGAAGCGCAGGGTAGACGCCCGCTG 1344
Db      343 ValHisSerMetGluTrpSerGlyGlnLysAspPheAlaLysThrAlaGluSerSerPhe 362
QY      1345 AAAGTCAACGGCGTCGAGTATGGGGAACACTCGCGAGTATGTAATTCCTCTCACTCGC 1404
Db      363 LeuValAspAspAlaGlnAlaGlyValLeuLysSerHisGlyAlaLeuSerPheLeuLys 382
QY      1405 GTCATGAGGCGAGCCATGAAGTCCCATCTACTACCAGCCCATCGCTCCCTGCAATGTTT 1464
Db      383 ValHisAsnAlaGlyHisMetValProMetAspGlnProLysAlaAlaLeuGluMetLeu 402
QY      1465 AACCGG 1470
Db      403 ArgArg 404

RESULT 13
A29412
C:carboxypeptidase C (EC 3.4.16.5) precursor - wheat
N:Alternate names: carboxypeptidase Y homolog
C:Species: Triticum aestivum (common wheat)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Nov-1999
C:Accession: A29412
R:Baulcombe, D.C.; Barker, R.F.; Jarvis, M.G.
J. Biol. Chem. 262, 13726-13735, 1987
A:Title: A gibberellin responsive wheat gene has homology to yeast carboxypeptidase Y.
A:Reference number: A29412; MUID:88007602; PMID:2820978
A:Accession: A29412
A:Molecule type: mRNA
A:Residues: 1-499 <BNA>
C:Genetics:
A:Introns: 100/2; 156/3; 196/3; 244/3; 295/1; 324/3; 411/1; 455/3
C:Superfamily: serine carboxypeptidase
C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase; zymogen
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-73/Domain: propeptide #status predicted <PRO>
F:74-484/Product: carboxypeptidase C #status predicted <MAT>
F:485-499/Domain: carboxyl-terminal propeptide #status predicted <CPRO>
F:144/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:216,404,461/Active site: Ser, Asp, His #status predicted

Alignment Scores:
Pred. No.:      3,47e-20      Length:      499
Score:          413.50      Matches:     126
Percent Similarity: 41.37%      Conservative: 80
Best Local Similarity: 25.30%      Mismatches: 217
Query Match:     14.24%      Indels:      75
DB:              1          Gaps:          16

US-09-712-338-1_COPY_55_1662 (1-1608) x A29412 (1-499)

QY      1 CTTCCAGGAGTACACCGCGCTCGTGGTAGAAGACAGCTACCCCAAGAACCCCGCG 60
Db      47 LeuProGlyArgProArgArgGlyLeuGly-----AlaGly 58
QY      61 GTCAGACTCTTACACCCCAACAAATGTCACATCCGTCGATACAGGAACCCGG---GCA 117
Db      59 AlaGluAspValAlaProGlyGlnLeuLeuGluAlaArgValThrLeuProGlyLeuPro 78
QY      118 GAGGCGTCTCGGAGACTACCCCGGTGTCAATCTCTACTCTGGATAT-----GTGAC 171
Db      79 GluGlyValGlyAsp-----LeuGlyHisHisAlaGlyTyrTrpArgLeuPro 94

```

```

QY      172 ACCCTCCCGAGTCCCATACCTCTCTCTGTTCTTCCGAAGCCAGACATAACCCAGAACT 231
Db      95 AsnThrHisAspAlaArgMetPheThrPhePhePheGluSerArgGlyLysLysGluAsp 114
QY      232 GCACCTATCACATGTGTTGAATGTCGCCCTCGGAAGCGATTTGTCGTCGTCCTTC 291
Db      115 ---ProValValIleTrpLeuThrGlyGlyProGlyCysSerSerGluLeuAlaValPhe 133
QY      292 GAAGAGTGGCGCTTGCATGTCATTCGACTTTCGATGATGATGATGATGATGATGATGAT 351
Db      134 TyrGluAsnGlyProPheThrIleAlaAsnMetSerLeuValTrpAsnLysPheGly 153
QY      352 TGGACAGAGGTCCTCAATTTACTATTCTCTGCCAGCATTCGGAGTGGCTTTTCATAT 411
Db      154 TrpAspLysIleSerAsnIlePheValAspProAlaThrGlyThrGlyPheSerIle 173
QY      412 AGTGATACGGTGTGATGGTTCATTAACCCCTGTAAGTGGGTGGTGGAAATTCGAGCTT 471
Db      174 SerSerAspAspArgAspThrArgHisAspGluAlaGlyVal----- 187
QY      472 GCAGGAGTTCAGGGCGGTACCCACCATTTGAGCCACTCTGATGATGATGATGATGATGAT 531
Db      187 ----- 187
QY      532 GCCGAGAGCGCGCTTGGGAGATCCTGCAAGATTCCTTAGTGGACTACCTAGCTTGGAC 591
Db      188 ---SerAsnAspLeuTyrAspPheLeuGlnValPhePheLysLysHisProGluPhe--- 205
QY      592 ICTAGGGTGCACCTAAGGACTTCAGTCTATGAGCGGAGAGCTATGGAGGCGACTATGTT 651
Db      206 -----ValLysAsnAspPheIleThrGlyGluSerIleThrGlyHisIle 222
QY      652 CCTGCATCTTCATTCATTTTACGACGAGATGAGAGATTCGCAACGAGTGCACACGGTAGT 711
Db      223 ProAlaPheAlaSerArgValHisGlnGlyAsnLysLys-----AsnGlu 237
QY      712 GGTCTTCAGCTTAATTTCAACTCTCTGGGATTTATACGGCATCATCGAGGCGGATC 771
Db      238 GlyThrHisIleAsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAspProAlaIle 257
QY      772 CAGSCCCTTACTACCCGATTCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 831
Db      258 GlnTyrLysAlaTyrThrAspTyrAlaLeuAsp-----MetAsnLeuIleGln 273
QY      832 GAGACCGCTCTACAACTACATGAAGTTTCCCAACCAATGCCAATGTTGCCAGGATTTG 891
Db      274 LysAlaAspTyrAspArgIle-----AsnLysPheIleProProCysGluPheAla 290
QY      892 ATTTCCACCTGCAACAGACAAACCCGC---ACCGCATTCAGTACGCGCTCTGCGGCC 948
Db      291 IleLysLeuCysGlyThrAspGlyLysAlaSerCysMetAlaAlaTyrMetValCysAsn 310
QY      949 GAAGCCACCAACATGTCAGGGACAATGTTGAGGGGCCATACACTAGGCTTTGCTGTCGT 1008
Db      311 SerIlePheAsnSerIleMetLysLeuValGlyThrLysAsnTyr----- 325
QY      1009 GGTGTGATGATATTCGGCATTCATATGATGACCGCATTCGCGCAAGTTATTACAC--- 1065
Db      326 -----TyrAspValArgLysGluCysGluGlyLysLeuCysTyrAspPheSerAsnLeu 343
QY      1065 ---AAATTCTGCCAAGGACTCTGTCTATGGAGCTATCGCGTC---AACATCAACTAC 1119
Db      344 GluLysPhePheGlyAspLysAlaValArgGlnAlaIleGlyValGlyAspIleGluPhe 363
QY      1120 ACCCAGTCCAAATACGCTACTACGCTTTCAGCAACAGCGGACITTTCTGTCGCC 1179
Db      364 ValSerCysSerThrSerValTyrGlnAlaMetLeuThrAsp-----TrpMet 379
QY      1180 AACTTCATCGAA---GACCTCGAGGAGATCTTGTCTCTCCCGTGGTGTCTCTCCCTCATC 1236
Db      380 ArgAsnLeuGluValGlyIleProAlaLeuLeuGluAspGlyIleAsnValLeuIleTyr 399

```





```

QY 1069 -----TTTCIG 1074
Db 1841 SerSerAspPheLeuAsnGlyTyrProCysTyrAspSerAlaMetGluGlyTyrLeu 1860
QY 1075 GCAAGGACTCTGTCATGGAGCTATCGCGTC-----ACATCAACIACACCCAGTCC 1128
Db 1861 AsnArgProValValArgLysAlaLeuAsnIleProAspSerValProTyrTyrAlaAla 1880
QY 1129 AATAATGACGTCTACTACGCTTTCCAGCAACAGCGCCTTTGTCIGGCCCACTTCAIC 1188
Db 1881 AsnAsnIleIleAsnAlaTyrAsnGlnValAspSerIleValProAsn----- 1898
QY 1189 GAAGACCTCGAGAGATCCTTGCT-----CTCCCGTGGGTCTCCCTCAICTAT 1239
Db 1899 -----LeuGlnIleIleMetAlaAsnAlaProAlaAsnPheLysMetLeuLeuTyrSer 1916
QY 1240 GCGAGCGCGGATTCATCTCAACTGTTCCGCGGTGAG----- 1278
Db 1917 GlyAspAlaAspThrMetValAsnTrpLeuGlyAlaGluIlePheThrAlaAsnAsnPhe 1936
QY 1279 ---GCCGTTCCTCGCTGCGAACTACTCCCAAGCCGCC-----CAGTTCCGAAGC 1326
Db 1937 AlaAlaLeuGlyLeuThrThrSerSerProArgAlaGlnTrpThrPheGlnTyrAsnSer 1956
QY 1327 GCAGGTACACGCGCCCTGAAGTCAACGGCTCGAGTATGGGAACTCGCGAGTATGCT 1386
Db 1957 Thr---PheGlnPro---ThrValAlaGlyTyrGlnThrSerTyrThrSerAsnAlaIle 1974
QY 1387 AATTCTCTCTCACTCGCTCTATGAGGAGGCGCATGATTCCTGGTGGGATATCGCAGAGGCCAGAAG 1506
Db 1975 AsnIleAspValLeuThrValLysGlySerGlyHisPheValProLeuAspArgProGln 1994
QY 1447 GCCTCCTCGAATGTTTAACCGGACTATCTTCGGTGGGATATCGCAGAGGCCAGAAG 1506
Db 1995 GlnAlaLeuGlnMet-----IleTyrAsnPheValLysSerArgGlyTyrAsn 2010
QY 1507 AAGATCTGCGCCAGCTACAGACGATGAAGCGGTACAGCTACGATACACAGTCTGCTC 1566
Db 2011 ThrProPheAspLeuAsnSerAsnPheThrThrThrThrThrThrThrThrThrThrThr 2030
QY 1567 GTGCG-----CTGCTACGCTACCAAGC 1590
Db 2031 ThrProGlyThrGlyProThrValThrAla 2040

RESULT 15
serine-type carboxypeptidase like protein - Arabidopsis thaliana
N:Alternate names: protein F4F15.110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C:Accession: T49079
R:Alcaraz, J.P.; Clabault, G.; Cottet, A.; Mache, R.; Mewes, H.W.; Rudd, S.; Lemcke, K.;
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25015
A:Accession: T49079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <ALC>
A:Cross-references: EMBL:AL049711; GSPDB:GN00061; ATSP:F4F15.110
A:Experimental source: cultivar Columbia; BAC clone F4F15
C:Genetics:
A:Gene: ATSP:F4F15.110
A:Map position: 3
A:Introns: 116/1; 151/1
C:Superfamily: serine carboxypeptidase

Alignment Scores:
Pred. No.: 4e-18 Length: 482
Score: 383.00 Matches: 135
Percent Similarity: 39.42% Conservative: 68
Best Local Similarity: 26.21% Mismatches: 202
Query Match: 13.19% Indels: 110

```

```

DB: 2 Gaps: 18
US-09-712-338-1_copy_55_1662 (1-1608) x T49079 (1-482)
QY 43 CCCAAGAACCCACCAGCGGTCAAGACTCTTACACCCGCAACAATGTCCATCCGGTAC 102
Db 32 ProSerProLysArgGlyValSerSerSerGlyAspThrSerHisPheAsnValIleGln 51
QY 103 AAGAA-----CCCGGGCGAGAGCGGTCTCGGAGACTACCCCGGGT----- 144
Db 52 ArgGluSerValProSerProLysAspLysLeuIleGlnLeuProGlyClnPro 71
QY 145 -----GTCAATCTCTACTCTGATATGTCACACTCTCCGAGTCCCATACC 192
Db 72 SerAspValThrPheLysGlnIyrGlyGlyTyrValAlaValAsnLysProAlaGlyArg 91
QY 193 TTC-----TTCCTGTTCTTCGAAGCCAGACATACCCAGAACTGCACCTATCACATTG 246
Db 92 PheLeuTyrTyrIyrPheValGluThrIleLysProGlyAsnThrThrProLeuValIle 111
QY 247 TGGTTGAAGTGGCCCTGGAAGCATTCCTTTGATCGGTCTCTTCGAAGAGTTCGGCCCI 306
Db 112 TrpPheAsnGlyGlyProGlyCysSerSerLeuGlyGlyAlaPheLysGluLeuGlyPro 131
QY 307 TGCATGTCAATTCGACTTTTGATGAC---TACATCAACCTCCTACTCGTGGAGAGGTC 363
Db 132 PheArgValHisSerAspGlyLysThrLeuPheArgAsnProTyrSerTrpAsnAsnGlu 151
QY 364 TCCAATTTACTATTCCTGTCGCCAGCATTCCTTTGATCGGTCTCTTCGAAGAGTTCGGCCCI 423
Db 152 AlaAsnValLeuPheLeuGluThrProValGlyThrGlyPheSerTyrSer----- 168
QY 424 GATGGTCCATTAACCTGTAACTGGGTCTCGAAAATTCGAGCTTTGAGGAGTTCAG 483
Db 169 -----AsnSerProIleAsnGlyLysGln 176
QY 484 GSCCGGTACCCACCACTTGATGCCACTCTGATCGACTACTACCAATCTTGCCGAGAGGCC 543
Db 177 GlyAspLysAlaThr-----AlaGluAsp 184
QY 544 GCTGGGAGATCCTCAAGGATTCCTTAGTGACTACCTAGCTTGACTCTAGGGTGCAG 603
Db 185 AsnTyrMetPheLeuValAsnTrpLeuGluArgPheProGluTyrLysGlyArg----- 202
QY 604 TCTAAGGACTTCAGTCTATGGAGGAGAGCTATGGAGGCACTATGCTCTCATCTTC 663
Db 203 -----AspIleTyrIleAlaGlyGlnSerTyrAlaGlyHisTyrValProGlnLeuAla 220
QY 664 AATCAITTTTACGACAGATGAGAGAATTGCCACGGTAGTGTAAIGTGTTCAGCTT 723
Db 221 GlnIleIleLeuHisArgAsn-----AsnGlnThrLeuIle 232
QY 724 AATTTCAACTCTCTGGGATTTATTACGGCATCATCGAGGCGGATCCAGGCCCTTAC 783
Db 233 AsnLeuArgGlyIleLeuIleGlyAsnProSerLeuAsnArgGluIleGlnAspPhe 252
QY 784 TACCCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACCGCTAC 843
Db 253 GlyTyrLysPheMetPhe-----SerHisGlyLeuIleSer---GlnGlnGlnMetAsp 269
QY 844 AACTACATGAAGTTTGCCCAACCAATGCCAAATGTTGCCAGGATTTGATTTCCACCTGC 903
Db 270 AsnTyrAsnLysPhe-----CysThrAsp-----SerAspLeu 280
QY 904 AACACAGACAAACCGCACCGCATTAAGTACTGACTACGCCCTCTCGCCGAGACCCACCAATG 963
Db 281 TyrAspTrpAspLysCysHisLeuAlaSerGlnLysIleGluAlaGlnLysThrHisLeu 300
QY 964 TGCAGGGACAATGTTGAGGGGCCACTACTACGCCCTTGTGCTGCTGTTGTTGTTATGATT 1023
Db 301 -----AspIleIyrAsnIle 305
QY 1024 CGGCATCCATAT----- 1035

```

wed Sep 17 15:23:15 2003

```

Db      306 TyrAlaProLeuCysLeuAsnSerThrLeuSerSerGluProLysLysCysThrThrIle 325
QY      1036 -----GATGACCGACTCCGCCAAGTATTACAAATAATTTCTGGGAAAGGACTCTGTC 1089
Db      326 MetLysAlaAspProCysSerGlyAsnTyrLeuLysAlaTyrLeuAsnIleLysGluVal 345
QY      1090 ATGACGGTATCGCGGTCAC-----ATCAACTACACCCAGTCCCAATAAT 1134
Db      346 GlnGluAlaIleHisAlaAsnThrThrLysIleProLysGluTrpThrSerCysAsnThr 365
QY      1135 GACGCTCTACTAGGCTTTCCAGCAACAGCGACTTTGCTGGCCCAACTTCATCGAAGAC 1194
Db      366 LysLeuLeuTrpGluTrpAsnGluLysAspArgTyrVal-----SerLeuThrProIle 383
QY      1195 CTCGAGGAGATCCTTCTCTCCCGTGGTGTCTCCCTCATCTATGGCGACGCCGATAC 1254
Db      384 LeuGlnGluLeuMetGlyLysGlyValArgValMetLeuTyrAsnGlyAspValAspLeu 403
QY      1255 ATCTGCAACTGGTTCGGCGGTTCAGCGCGTTTCCCTCGCTGGGAACTACTCCCAAGCCGCC 1314
Db      404 ValIleProPheThrSerThrLeuAlaValValLysThrMetAsnLeuThrValValLys 423
QY      1315 CAGTTCCGGAAGCGCAGGTACACGCCCTGAAAGTCAACGCGCTCGAGTATGGGAAACT 1374
Db      424 GluTrpArg-----ProTrpPheThrGlyGlyHisValGlyGlyPheThr 438
QY      1375 CGCGAGTAT---GGTAATTCTCCTTCATCTCGCTCTATGAGGAGCGCCATGAAGTCCCA 1431
Db      439 GluAspTyrLysGlyAsnLeuThrPheValThrValLysGlyAlaGlyHisSerValPro 458
QY      1432 TACTACCGCCATCGCCCTCCCTGCAATTGTTTAAACCGGACTATC 1476
Db      459 ThrAspGlnProIleHisAlaLeuAsnIlePheThrSerPheIle 473

```

Search completed: September 16, 2003, 18:48:40  
Job time : 71 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 16, 2003, 18:04:58 ; Search time 29 Seconds

(without alignments)  
5215.099 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662

Perfect score: 2904

Sequence: 1 cttccaggaagtagcaccggc.....gcattgccaagtgtgtgtatg 1608

Scoring table:

	Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US09712338/runat\_16092003\_144325\_14345/app\_query.fasta\_1.1799  
-DB=SwissProt\_41 -QPMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=ptc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFM=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09712338@cgn\_1\_1\_24@runat\_16092003\_144325\_14345 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	894	30.8	423	1	CPSL_PENJA
2	652.5	22.5	1002	1	CPFY_SCHPO
3	636	21.9	508	1	YBY9_YEAST
4	599	20.6	542	1	CBPY_CANAL
5	594.5	20.5	523	1	CBPY_PICPA
6	527.5	18.2	532	1	CBPY_YEAST
7	474	16.3	516	1	CBPX_ARATH
8	473	16.3	429	1	CBPX_ORYSA
9	452	15.6	500	1	CBP3_ORYSA
10	434	14.9	508	1	CBP3_HORVU
11	413.5	14.2	500	1	CBP3_WHEAT
12	359	12.4	476	1	CPVL_HUMAN
13	353.5	12.2	436	1	CP22_HUMAN
14	352.5	12.1	470	1	YSS2_CAEEL
15	351	12.1	482	1	NF31_NAEFO
16	349	12.0	474	1	PRTP_MOUSE
17	344.5	11.9	516	1	CP23_HORVU
18	332	11.4	469	1	YOW5_CAEEL

19	329.5	11.3	480	1	PRTP_HUMAN	P10619	homo sapien
20	326.5	11.2	454	1	YUA6_CAEEL	P52715	caenorhabdi
21	321.5	11.1	729	1	KEX1_YEAST	P09620	saccharomyc
22	318.5	11.0	476	1	CBP2_HORVU	P08818	hordeum vul
23	316	10.9	574	1	YBP3_CAEEL	P52716	caenorhabdi
24	312.5	10.8	471	1	VCP_AEADAE	P42660	aedes aegypt
25	312	10.7	510	1	CBP1_ORYSA	P37890	oryza sativ
26	305.5	10.5	423	1	CBP2_WHEAT	P08819	tritium ae
27	297.5	10.2	505	1	YXD2_CAEEL	P52714	caenorhabdi
28	295.5	10.2	286	1	CBPX_PEA	Q41005	pisum sativ
29	280	9.6	499	1	CBP1_HORVU	P07519	hordeum vul
30	269.5	9.3	523	1	PEPS_ASPSA	P52719	aspergillus
31	255	8.8	452	1	RISC_MOUSE	Q920a5	mus musculus
32	248	8.5	452	1	RISC_RAT	Q920a6	rattus norv
33	237	8.2	452	1	RISC_HUMAN	Q9hb40	homo sapien
34	225.5	7.8	507	1	SXA2_SCHPO	P32825	schizosacch
35	221.5	7.6	531	1	PEPF_ASPNG	P52718	aspergillus
36	206	7.1	5703	1	MUSB_HUMAN	Q9hc84	homo sapien
37	195.5	6.7	5179	1	MUC2_HUMAN	P47179	saccharomyc
38	191.5	6.6	1161	1	DAN4_YEAST	P08640	saccharomyc
39	190	6.5	1367	1	AMYH_YEAST	P55747	hordeum vul
40	167.5	5.8	324	1	CP21_HORVU	P38088	homo sapien
41	166.5	5.7	1233	1	MU5A_HUMAN	P03200	epstein-bar
42	166	5.7	907	1	VGP3_EBV	P32323	saccharomyc
43	160	5.5	725	1	AGAL_YEAST	Q99102	homo sapien
44	159	5.5	610	1	MUC4_HUMAN	P28968	equine herp
45	157	5.4	797	1	VGLX_RSVEB		

## ALIGNMENTS

RESULT 1  
CPSL\_PENJA  
ID CPSL\_PENJA STANDARD; PRT; 423 AA.  
AC P34946;  
DT 01-FEB-1994 (Rel. 28, Created)  
DI 01-FEB-1994 (Rel. 28, Last sequence update)  
DI 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Carboxypeptidase SI (EC 3.4.16.6).  
OS Penicillium janthinellum (Penicillium vitale).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.  
OX NCBI\_TaxID=5079;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=94039747; PubMed=8224168;  
RA Svendsen I., Hofmann T., Endrizzi J., Remington S.J., Breddam K.;  
RT "The primary structure of carboxypeptidase SI from Penicillium  
janthinellum".  
RL FEBS Lett. 333:39-43(1993).  
CC -!- CATALYTIC ACTIVITY: Preferential release of a C-terminal arginine  
or lysine residue.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.  
DR PIR: S38953; S38953.  
DR HSSP: P00729; IYSC.  
DR MEROPS: S10.008; -.  
DR InterPro: IPR000379; Ser\_estrs\_site.  
DR InterPro: IPR001563; Serine\_carbpept.  
DR Pfam: PF00450; serine\_carbpept; 1.  
DR PRINTS: PR00724; CBBCXYPIASEC.  
DR ProDom: PD001189; Serine\_carbpept; 1.  
DR PROSITE: PS00131; CARBOXYPEPT\_SER\_SER; 1.  
DR PROSITE: PS00560; CARBOXYPEPT\_SER\_HIS; 1.  
KW Hydrolase; Carboxypeptidase; Glycoprotein.  
FT ACT\_SITE 143 143  
FT ACT\_SITE 340 340 BY SIMILARITY.  
FT ACT\_SITE 397 397  
FT BINDING 343 343 SUBSTRATE.  
FT BINDING 398 398 SUBSTRATE.  
FT DISULFID 8 68  
FT DISULFID 55 300  
FT DISULFID 223 246  
FT DISULFID 230 239

FT CARBOHYD 200 200 N-LINKED (GLCNAC...).  
SQ SEQUENCE 423 AA; 46496 MW; 55F3333608B12729 CRC64;

Alignment Scores:  
Pred. No.: 1,02e-54 Length: 423  
Score: 894.00 Matches: 186  
Percent Similarity: 54.31% Conservative: 66  
Best Local Similarity: 40.09% Mismatches: 160  
Query Match: 30.79% Indels: 52  
DB: 1 Gaps: 8

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x CPS1\_PENJA (1-423)

1018 GATATTGGCATCCATATGATGACCGGAGTTCGCGCAAGTATTATACACAAATTTCTGGCA 1077  
|||||  
264 AspIleArgGluProSerAsnAspProPyrProPylsThrTyrSerThrTyrLeuSer 283  
|||||  
1078 AAGGACTCTGTATGGAGCGCTATCGCGCTCAACATCAACTACACCGAGTCCCAATATGAC 1137  
|||||  
284 AspProThrValValValValValValValValValValValValValValValValVal 303  
|||||  
1138 GTCTACTAGCTTCCAGCAAGGAGGACTTTGTCTGCGCCCAACTCATCGAAGACCTC 1197  
|||||  
304 ProTyrAsnLysPheAlaSerThrGlyAspAsnPro---ArgSerPheLeuSerThrLeu 322  
|||||  
1198 GAGGAGATCTTCCTCTCCCGTGGTGTCTCCCTCATCTATGCGGACCGCCGATTACATC 1257  
|||||  
323 SerSerValValValSerGlyIleAsnValValValValValValValValValValVal 342  
|||||  
1258 TCAACTGTGTGGCGGTGTCAGGCGCTTCCCTGCGCGCACTACCTCCGACGCCGCCAG 1317  
|||||  
343 CysAsnTrpLeuGlyAsnTyrGluValAlaAsnAlaValAspPheProGlyAsnAlaGln 362  
|||||  
1318 TTCCGAGCGCGAGGTACAGCGCCCTCAAGTCAAGCGGTGCGAGTATGGGAAATCGC 1377  
|||||  
363 PheSerAlaLeuAspLeuAlaProTyrThrValAsnGlyValGlnLysGlyGlnPheLys 382  
|||||  
1378 GAGTATGGTAATTCCTCTCATCTGCTATGAGCGAGCGCCGATCAAGTCCCAATAC 1437  
|||||  
383 ThrValAspAsnPheSerPheLeuLysValTyrGlyAlaGlyHisGluValProTyrTyr 402  
|||||  
1438 CAGCCATCGCCCTCCCTGCAATTTTAAACCGCACTATCTCGGTGGGATATCGCAGAG 1497  
|||||  
403 GlnProAspThrAlaLeuGlnAlaPheLysGlnIleLe 415  
|||||  
1498 GGCCAGAGAAG 1509  
|||||  
416 ---GlnLysLys 418

RESULT 2  
CBPX\_SCHPO STANDARD; PRT: 1002 AA.  
ID CBPX\_SCHPO  
AC 013849; 014366;  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Carboxypeptidase Y precursor (EC 3.4.16.5) (CPY).  
GN CPY1 OR CPY1 OR SPAC19G12.10C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A., SUBUNIT, SUBCELLULAR LOCATION, AND MUTAGENESIS OF  
CYS-627.  
RX MEDLINE=97352672; PubMed=9209031;  
RA Tabuchi M., Iwahara O., Ohtani Y., Ohuchi N., Sakurai J.-I.,  
Morita T., Iwahara S., Takegawa K.;  
RT "Vacuolar protein sorting in fission yeast: cloning, biosynthesis,  
transport, and processing of carboxypeptidase Y from  
Schizosaccharomyces pombe.";  
RL J. Bacteriol. 179:4179-4189(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
Brooks J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
Collins M., Connor R., Cronin A., Davis P., Hildalgo J., Hodgson G.,  
Gentles S., Goble A., Hamlin N., Harris D., Huckle E.J., Hunt S., Jagels K.,  
Hollroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,

121 GCGCTCTGGAGACTACCCGGGTCAAACTCTACTCTGGATATGTCGACACCTCTCC 180  
6 GlyIleCysGluThrProGlyValAsnGlnTyrSerGlyTyrLeuSerValGlySer 25  
181 GAGTCCCATACCTCTCTGTTCTTCGAAGCCAGACATACCCAGAACTCCACTATC 240  
26 AsnMetAsnMetTrpPheTrpPheGluAlaArgAsnAsnProGlnGlnAlaProLeu 45  
241 ACATTTGGTGTGAATGGTGGCCCTGGAACGGAATCTTTGATCGGCTCTTCCAGAGTTG 300  
46 AlaAlaTrpPheAsnGlyGlyProGlyCysSerMetIleGlyLeuPheGlnGluAsn 65  
301 GGCCTTGGCAT---GTCAATTCGACTTTTGATGACTACATCAACCTCCTCGTGGAA 357  
66 GlyProCysHisPheValAsnGlyAspSerThrProSerLeuAsnGluAsnSerTrpAsn 85  
358 GAGGTCTCAATTTACTATCTCTGCCAGCATTTGGAGTGGCGCTTTTCATATAGTAT 417  
86 AsnTyrAlaAsnMetIleTyrIleAspGlnProIleGlyValGlyPheSerTyrGly--- 104  
418 ACGTTGATGGTGGTCCATTACCTGTAACCTGGGTGCTCGAAATTCGAGCTTGCAGGA 477  
104 -----  
478 GTTCAGGCGGTACCAACCACTGATGCCACTCTGATCGATATACCAATCTTCCGCA 537  
105 -----ThrAspAspValThr-----SerThrValThrAlaAla 115  
538 GAGGCGCTTGGAGATCTCTGGAATTCCTTAGGAGTACTACTAGCTGGACTCAGG 597  
116 ProTyrValTrpAsnLeuGlnAlaPheTyrAlaGlnArgProGluTyrGluSerArg 135  
598 GTGCAGCTTAGGACTTCACTGATGACGAGAGCTATGAGGCACTATGTCCTGCA 657  
136 -----AspPheAlaIlePheThrGluSerTyrGlyGlyHisTyrGlyProGlu 151  
658 TTCTTCATCATTTTACGAGCAGATGACAGAAATTCGCAACGCTAGTGTAAATGGTGT 717  
152 PheAlaSerTyrIleGluGlnGlnAsnAlaAlaIleLysAlaGlySerValThrGlyGln 171  
718 CAGCTTAATTTCAACTCTCTGGGAATTTATTAACGGCATATCGACGAGCGCATCCAGCC 777  
172 AsnValAsnIleValAlaLeuGlyValAsnAsnGlyTrpIleAspSerThrIleGlnGlu 191  
778 CCTTACTACCTGAACTGCTGTGACAAATACCTACGCTATCAAGGCTGTCAAGCAGACC 837  
192 LysAlaTyrIleAspPheSerTyrAsnAsnSerTyrGlnGlnIleLeuAspSerThr 211  
838 GTCTACACTACATCAAGTTTCCCAACCAATGTCGAGGATTTGATTTGATTTCC 897  
212 ArgAspSerLeuLeuAspAlaTyrAsn-----AsnGlnCysLeuProAlaLeuGln 228  
898 ACCTGCAACAGACAAACCGCACCGCATTAGCTGACTAGCCCTCTGCGCGCAACCCACC 957  
229 GlnCysSerGlnSerGlySerThr-----SerAspCysThrAsnAlaAsp 243  
958 ACATGTCGAGGACAAATCTTTCAGGGCCATACAGCTTTGCTGGTGGTGGTGTAT 1017  
244 SerValCysTyrGlnAsnIleGluGlyProIleSerSerSerGlyAspPheAspValTyr 263

FT	415	423	2-7 (APPROXIMATE).
REPEAT	715	715	BY SIMILARITY.
FT ACT_SITE	921	921	BY SIMILARITY.
FT ACT_SITE	978	978	BY SIMILARITY.
FT BINDING	924	924	BY SIMILARITY.
FT BINDING	979	979	SUBSTRATE (BY SIMILARITY).
FT DISULFID	627	880	SUBSTRATE (BY SIMILARITY).
FT DISULFID	776	789	BY SIMILARITY.
FT DISULFID	799	822	BY SIMILARITY.
FT DISULFID	806	815	BY SIMILARITY.
FT DISULFID	844	851	BY SIMILARITY.
FT CARBOHYD	659	659	BY SIMILARITY.
FT MUTAGEN	627	627	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT			C->T: 36% OF ORIGINAL ACTIVITY.
SO	SEQUENCE	114237	AA; 114237 MW; 4A8D81CFDAB2D854 CRC64;

  

Alignment Scores:	
Pred. No.:	8,9e-38
Score:	652.50
Percent Similarity:	47.34%
Best Local Similarity:	33.81%
Query Match:	22.47%
DB:	1

  

US-09-712-338-1_COPY55_1662 (1-1608) x CBPY_SCHPO (1-1002)	
QY	91 ACATCCGGTACAAAGAACCCGGGCGAGGGCGCTCTCGAGACTACCCCGGGTGTCAA 150
DB	570 ThrLeuArgValLysAspSerLysProGluSerLeuGlyIleAspThr-----ValLys 587
QY	151 TCCTACTCTGGATGTCGACACCTCTCCCGAGTCCCATACCTTCTTCTGTTCTTCGAA 210
DB	588 GluTyrThrGlyTyrLeuAspValGluAspAspArgHisLeuPhePheThrPheGlu 607
QY	211 GCCACATACCCGAACCTGCACCTATACATTTGGTTGATGGTGGCCCTGGGAAGC 270
DB	608 SerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsnGlyGlyProGlyCys 627
QY	271 GATTCCTTTGATCGGTCCTCTCGAAGAGTGGCCCTTGCATGTCAT---TCGACTTTT 327
DB	628 SerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIleAsnIleGluThrLeu 647
QY	328 GATGACTACATCAACCCCTCACICGTGGAGAGGCTCTCCAAATTTACTTTCGTGCCAG 387
DB	648 LysProGluTyrAsnProHisSerTrpAsnSerAsnAlaSerValIlePheLeuAspGln 667
QY	388 CCATTGGGAGTCGGCTTTTCATATAGTCATACGGTTGATGGGTCCATTAAACCTGTAACT 447
DB	668 ProIleAsnThrGlyPheSerAsnGlyAsp-----AspSerValLeuAspThrValThr 685
QY	448 GGGGTCTGCGAAATTCGAGCTTTGCAGGAGTTTCAGGGCCGGTACCACCAATTGATGCC 507
DB	685 ----- 685
QY	508 ACTGTGATCGATACTACCAATCTTGCCGAGAGGCGGCTTTGGAGAGTCTGCAAGGATTC 567
DB	686 -----AlaGlyLysAspValTyrAlaPheLeuAsnLeuPhe 697
QY	568 CTTAGTGGACTACTAGCTGGGACTCTAGGGTCAGTCTPAGGACTTCAGTCTATGGAGC 627
DB	698 PheAlaLysPhePro-----GluTyrAlaHisLeuAspPheHisIleAlaGly 713
QY	628 GAGAGCTATCGAGGGCACTATGGTCIGCATCTTCTCAATCATTTTTTACGACGAGATGAG 687
DB	714 GluSerTyrAlaGlyHisTyrIleProGluPheAlaLysGluIleMetGluHisAsnGln 733
QY	688 AGA-----ATTGCCAACGGT---ACTGGTTAATGGTGTTGTCACCTTAATTTCAC 732
DB	734 GlyAlaAsnPhePheValAlaSerGlyTyrGluMetGluLysGlnTyrIleAsnLeuLys 753
QY	733 ICTCTGGGAATTAITAAACGGCATATCGACGAGGCGATCCAGGGCCCTTACTACCCCTGAA 792
DB	754 SerValLeuIleGlyAsnGlyLeuThrAspProLeuValGlnTyrThrPheMetGluHis 772

FT	415	423	2-7 (APPROXIMATE).
REPEAT	715	715	BY SIMILARITY.
FT ACT_SITE	921	921	BY SIMILARITY.
FT ACT_SITE	978	978	BY SIMILARITY.
FT BINDING	924	924	BY SIMILARITY.
FT BINDING	979	979	SUBSTRATE (BY SIMILARITY).
FT DISULFID	627	880	SUBSTRATE (BY SIMILARITY).
FT DISULFID	776	789	BY SIMILARITY.
FT DISULFID	799	822	BY SIMILARITY.
FT DISULFID	806	815	BY SIMILARITY.
FT DISULFID	844	851	BY SIMILARITY.
FT CARBOHYD	659	659	BY SIMILARITY.
FT MUTAGEN	627	627	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT			C->T: 36% OF ORIGINAL ACTIVITY.
SO	SEQUENCE	114237 MW; 114237 MW; 4A8D81CFDAB2D854 CRC64;	

  

Alignment Scores:	
Pred. No.:	8,9e-38
Score:	652.50
Percent Similarity:	47.34%
Best Local Similarity:	33.81%
Query Match:	22.47%
DB:	1

  

US-09-712-338-1_COPY55_1662 (1-1608) x CBPY_SCHPO (1-1002)	
QY	91 ACATCCGGTACAAAGAACCCGGGCGAGGGCGCTCTCGAGACTACCCCGGGTGTCAA 150
DB	570 ThrLeuArgValLysAspSerLysProGluSerLeuGlyLeuAspThr-----ValLys 587
QY	151 TCCTACTCTGGATGTCGACACCTCTCCCGAGTCCCATACCTTCTTCTGTTCTTCGAA 210
DB	588 GluTyrThrGlyTyrLeuAspValGluAspAspArgHisLeuPhePheThrPheGlu 607
QY	211 GCCACATACCCGAACCTGCACCTATACATTTGGTTGAATGGTGGCCCTGGGAAGC 270
DB	608 SerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsnGlyGlyProGlyCys 627
QY	271 GATTCCTTTGATCGGTCCTCTCGAAGAGTGGCCCTTGCATGTCAT---TCGACTTTT 327
DB	628 SerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerLeuAsnIleGluThrLeu 647
QY	328 GATGACTACATCAACCCCTCACICGTGGAGAGGCTCCCAATTTACTTTCGTGCCAG 387
DB	648 LysProGluTyrAsnProHisSerTrpAsnSerAsnAlaSerValIlePheLeuAspGln 667
QY	388 CCATTGGGAGTCGCTTTTCATATAGTCATACGGTTGATGGGTCCATTAAACCTGTAACT 447
DB	668 ProIleAsnThrGlyPheSerAsnGlyAsp-----AspSerValLeuAspThrValThr 685
QY	448 GGGGTCTGCGAAATTCGAGCTTTGCAGGAGTTTCAGGGCCGGTACCACCAATTGATGCC 507
DB	685 ----- 685
QY	508 ACTGTGATCGATACTACCAATCTTGCCGAGAGGCGCTTTGGAGATCCTGCAAGGATTC 567
DB	686 -----AlaGlyLysAspValTyrAlaPheLeuAsnLeuPhe 697
QY	568 CTTAGTGGACTACTAGCTGGGACTCTAGGGTCAGTCTAAGGACTTCAGTCTATGGAGC 627
DB	698 PheAlaLysPhePro-----GluTyrAlaHisLeuAspPheHisIleAlaGly 713
QY	628 GAGAGCTATGGAGGCACTATGGTCIGCATCTTCTCAATCATTTTTTACGACGAGATGAG 687
DB	714 GluSerTyrAlaGlyHisTyrIleProGluPheAlaLysGluIleMetGluHisAsnGln 733
QY	688 AGA-----ATTGCCAACGGT---ACTGGTTAATGGTGTTGTCACCTTAATTTCAC 732
DB	734 GlyAlaAsnPhePheValAlaSerGlyTyrGluMetGluLysGlnTyrIleAsnLeuLys 753
QY	733 ICTCTGGGAATTAITAAACGGCATATCGACGAGGCGATCCAGGCGCCCTTACTACCCCTGAA 792
DB	754 SerValLeuIleGlyAsnGlyLeuThrAspProIleValGlnTyrThrPheMetGluHis 772

QY 793 TTCGCTGTGGAACAATACCTACGGT-----ATCAAG 822  
 Db |||||  
 QY 774 MetAlaCysGlySerProTyrGlyProIleMetSerGlnGluGluCysAspArgIleThr 793  
 Db |||||  
 QY 823 GCTGTCAAGAGACCGCTCTACAACTACATGAAGTTGCGAACCAATGCAATGTTGTC 882  
 Db |||||  
 QY 794 GlyAlaTyrAspThr-----Cys 799  
 Db |||||  
 QY 883 CAGGATTTGATTTCCACCTGCAACAGACAAACCCGCGCATTAAGTACATGACATGCGCCTC 942  
 Db |||||  
 QY 800 AlaLysLeuIleThrGlyCysGlnThrGlyPheThr-----ProVal 814  
 Db |||||  
 QY 943 TCGCGCGAAGCCACCAACATCTGCGAGGACAAATGTTGAGGGCGCATACATACGCTTGTCT 1002  
 Db |||||  
 QY 815 CysIleGlyAlaSerLeuTyrCysAsnAsnAlaMetIleGlyPro---PheThrLysThr 833  
 Db |||||  
 QY 1003 GPCGTGGTGTATGATATGTCGATCCATATGATGACCGGACTCGCGCAAGTTAT--- 1059  
 Db |||||  
 QY 834 GlyLeuAsnIleTyrAspIleArgGluGluCysArgAspGlnGluHisLeuCysTyrPro 853  
 Db |||||  
 QY 1060 -----TACACAAATTTCTGCGAAGGACTCTGTCATGAGCGTATCGCGGTC 1107  
 Db |||||  
 QY 854 GluThrGlyAlaIleGluSerTyrLeuAsnGlnGluPheValGlnGluAlaLeuGlyVal 873  
 Db |||||  
 QY 1108 AACATCAACTACACCCAGTCCATATGATGCTTACTACGCTTTCCAGCAACAGCGGAC 1167  
 Db |||||  
 QY 874 GluTyrAspTyrLysGlyCysAsnThrGluValAsnIleGlyPheLeuPheLysGlyAsp 893  
 Db |||||  
 QY 1168 TTTGTCTGCGCCAACTTCATGCAAGACTCGAGAGATCCTT-----GCTCTCCCGTG 1221  
 Db |||||  
 QY 894 TrpMetArgLysThrPheArgAspValThrAlaIleLeuGluAlaGlyLeuProVal 913  
 Db |||||  
 QY 1222 CGTGTCTCCCTCATCTAT---GGCGAGCGGATTCATGCAAGTGGTTCGGGGGCGAG 1278  
 Db |||||  
 QY 914 -----LeuIleTyrAlaGlyAspAlaAspTyrIleCysAsnTyrMetGlyAsnGlu 930  
 Db |||||  
 QY 1279 GCGGTTTCCCTGCTGCGAACTACTCCAGCCCGCAGTTCGGAAGCGGAGGTATACG 1338  
 Db |||||  
 QY 931 AlaTrpThrAspAlaLeuGluTyrPalaGlyGlnArgGluPheTyrGluAlaLeuLys 950  
 Db |||||  
 QY 1339 CCCGTGAAGTCAACGCGGTGATGATGGGAAACTCCGAGTATGTTAAATTTCTCCTTC 1398  
 Db |||||  
 QY 951 ProTrpSerProAsnGlyLysGluAlaGlyArgGlyLysSerPheLysAsnPheGlyTyr 970  
 Db |||||  
 QY 1399 ACTCGGTCTATGAGGAGGATCACTCCATCCATACAGCCATCGCTCCCTCGAA 1458  
 Db |||||  
 QY 971 LeuArgLeuTyrGluAlaGlyHisMetValProPheAsnGlnProGluAlaSerLeuGlu 990  
 Db |||||  
 QY 1459 TTGTTTAAACGCGACTATCTCGGT 1482  
 Db |||||  
 QY 991 MetLeuAsnSerTrpIleAspGly 998  
 Db |||||  
 RESULT 3  
 ID YBY9\_YEAST STANDARD; PRT; 508 AA.  
 AC P38109.  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative serine carboxypeptidase in BSRI-IRAI intergenic region  
 DE (EC 3.4.16.-)  
 GN YBR139W OR YBR1015.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=288c;  
 RX MEDLINE=9437817; PubMed=8091856;  
 RA Becam A.-M., Cullin C., Grzybowska E., Lacroite F., Nasr F.,  
 RA Ozier-Kalogeropoulos O., Palucha A., Slonimski P.P., Zagulski M.,  
 RA Herbert C.J.;

"The sequence of 29.7 kb from the right arm of chromosome II reveals  
 13 complete open reading frames, of which ten correspond to new  
 genes.";  
 Yeast 10:S1-S11(1994).  
 [2]  
 RP DISCUSSION OF SEQUENCE.  
 RX MEDLINE=95042830; PubMed=7954890;  
 RA Nasr F., Becam A.-M., Grzybowska E., Zagulski M., Slonimski P.P.,  
 RA Herbert C.J.;  
 RT 'An analysis of the sequence of part of the right arm of chromosome  
 II of S. cerevisiae reveals new genes encoding an amino-acid permease  
 and a carboxypeptidase.';  
 RT Cur. Genet. 26:1-7(1994).  
 RL -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; X75891; CAA53497.1; -;  
 DR EMBL; Z36008; CAA85097.1; -;  
 DR PIR; S46008; S46008.  
 DR HSP; P00729; ICPY.  
 DR MEROPS; S10.0PW; -;  
 DR SGD; S0000343; YBR139W.  
 DR InterPro; IPR001379; Ser\_estrs\_site.  
 DR InterPro; IPR001563; Serine\_carbpept.  
 DR Pfam; PF00450; serine\_carbpept; 1.  
 DR PRINTS; PR00724; CRBOXYPTASEC.  
 DR ProDom; PD001189; Serine\_carbpept; 1.  
 DR PROSITE; PS00131; CARBOXYPEPT\_SER\_HIS; 1.  
 DR PROSITE; PS00560; CARBOXYPEPT\_SER\_HIS; 1.  
 KW Hypothetical protein; Hydrolase; Carboxypeptidase.  
 FT ACT\_SITE 219 219 BY SIMILARITY.  
 FT ACT\_SITE 415 415 BY SIMILARITY.  
 FT ACT\_SITE 474 474 BY SIMILARITY.  
 SQ SEQUENCE 508 AA: 57639 MW; AAB2806C8EE2EDE1 CRC64;  
 Alignment Scores: 1.03e-36 Length: 508  
 Pred. No.: 636.00 Matches: 167  
 Score: 46.25% Conservative: 70  
 Percent Similarity: 32.62% Mismatches: 197  
 Best Local Similarity: 21.90% Indels: 78  
 Query Match: 1 Gaps: 14  
 DB:

US-09-712-338-1-copy\_55\_1662 (1-1608) x YBY9\_YEAST (1-508)  
 QY 1 CTTCAGGAAGTACACCGGCTCGTCCGTAGACAGACGTA-----CCCAAG 48  
 Db |||||  
 QY 43 LeuProGlnAsnThrGlnGlnThrLeuLysLeuAspArgLeuAsnHisAspProLeu 62  
 Db |||||  
 QY 49 AACCCACCGGGGTCAAGACTCTTACACCGCAACAATGTCCACCTCGGTACAGGAA 108  
 Db |||||  
 QY 63 PheThrThrPheIleSerSerValAspThrAspTyrSerLeuArgLeuArgThrValAsp 82  
 Db |||||  
 QY 109 CCCGGGCGAGGGGCTGCGAGACTACCCCGGGTGTCAATCTCTCAATCTCTGATATGTC 168  
 Db |||||  
 QY 83 ProSerLysLeuGlyIle-----AspThrValLysGlnTrpSerGlyTyrMet 98  
 Db |||||  
 QY 169 GACACCTCTCCGAGTCCCATACCTCTCTGTTGGTTCTTCGAGCCAGACATACCCAGAA 228  
 Db |||||  
 QY 99 AspTyrLysAspSerLysHisPheIleTyrPhePheGluSerArgAsnAspProAla 118  
 Db |||||  
 QY 229 ACTGCACCTATCATATTCGTTGATGTTGGTGGCCCTGGAGCGGATTTCTTGTATCGGTCTC 288  
 Db |||||  
 QY 119 AsnAspProIleLeuLeuTrpLeuAsnGlyGlyProGlyCysSerSerPheThrGlyLeu 138  
 Db |||||  
 QY 289 TTCGAAGAGTGGGCGCTTCCCATGTGCAATTCGACTTTTGTATGACTACATCAACCTCAC 348  
 Db |||||

Db 139 LeuPheGluLeuGlyProSerSerIleGlyAlaAspMetLysProIleHisAsnProTyr 158  
 QY 349 TCGTGGACAGGAGTCCCAATTTACTATTCTCTCCAGCCATCGGAGTGGCGCTTTTCA 408  
 Db 159 SerTrpAsnAsnAlaSerMetIlePheLeuGluGlnProLeuGlyValGlyPheSer 178  
 QY 409 TATAGTATAGGTTGATGGGTCCATTAAACCTGTAACTGGGGTCGTCGAAATTCGAGC 468  
 Db 179 TyrGlyAsp 181  
 QY 469 TTTCAGGAGTTCAGGCGCGGTACCCAAACCATTCGATGCCACTCGATCGATACCTACCAAT 528  
 Db 182 188  
 QY 529 CTTGCCGAGAGCGCGCTTGGAGATCTCTGCAAGGATTCTTGTAGTACCTACCTAGCTG 588  
 Db 189 LeuAlaGlyLysAspAlaTyrIlePheLeuGluLeuPheGluAlaPheProHisLeu 208  
 QY 589 GACTCTAGGTGAGTCTAAGGACTTCAGTCTATGGACGGAGAGTATGGAGGCACTAT 648  
 Db 209 648  
 QY 649 GGTCTGCACTTCCTCAATCATTTTACGACCAAT---GAGAGAAATGCCAGGTA 705  
 Db 225 IleProGlnIleAlaHisGluIleValValLysAsnProGluArg 239  
 QY 706 GTTAATGGTTCAGCTTAATTCCTCACTCTCTGGGAATTAATACCGCATCATCGAGCAG 765  
 Db 240 765  
 QY 766 GCGATCCAGCGCCCTTACTACCTCGAATTCGCTGTG---AACAAATACCTACGATCAAG 822  
 Db 256 LeuIleGlnAlaAspTyrTrpGluProMetAlaCysGlyLysGlyTyrHisProVal 275  
 QY 823 CGTGTCAAGGACGCGTCTACACTACATCAAGTTTGGCAACCAATGCCAAATGTTGC 882  
 Db 276 LeuSerSerGluGluCysGluLysMetSerLysAlaAlaGlyArg 290  
 QY 883 CAGGATTTGATTCCACCTGCAACACAGACACCGCGCATTCAGTACGACGCTC 942  
 Db 291 942  
 QY 943 ---CysArgArgLeuAsnLysLeuCysTyrAlaSerLysSer 304  
 Db 305 LeuProCysIleValAlaThrAlaTyrCysAspSerAlaLeuLeuGluProTyrIleAsn 324  
 QY 997 TTTGCTGTCGTGGTGTATCATATTCGGCATCATATGATGACCCG 1044  
 Db 325 1044  
 QY 1045 ---ACTCGCCCAAGTTTATACAAACAATTCCTGGCAAGGACTCTCTCATGGAC 1095  
 Db 344 MetCysTyrThrGlyLeuArgTyrValAspGlnTyrMetAsnPheProGluValGlnGlu 363  
 QY 1096 GCTATCGCGTCAACATC---AATACACCCAGTCCAAATATGACCTACTACGCTTTC 1152  
 Db 364 ThrLeuGlySerAspValHisAsnTyrSerGlyCysAspAsnAspValPheThrGlyPhe 383  
 QY 1153 CAGAAACAGCGACTTTGTCTGGCCCACTTCATCGAAGACCTCGAGAGATTCCTGCT 1212  
 Db 384 LeuPheThrGlyAspGlySerLysPro---PheGlnGlnTyrIleAlaGluLeuLeuAsn 402  
 QY 1213 CTCCCGTGGCTGCTCTCCCTCATCTATGGCAGCGGATTCATCTGCAACTGGTTCGGC 1272  
 Db 403 HisAsnIleProValLeuIleTyrAlaGlyAspLysAspTyrIleCysAsnIleLeuGly 422  
 QY 1273 GGTACGCGCTTCTCCTCGTGGCACTACTCCCAAGCGCGCGAGTCCGAGCGCAGG 1332  
 Db 423 AsnHisAlaTrpSerAsnGluLeuGluTrpIleAsnLysArgArgTyrGlnArgMet 442  
 QY 1333 TACAGGCC-----CTGAAAGTCAACGGCTCGAGTATGGGAACCTCGCGAGTATG 1386

Db 443 LeuArgProTrpValSerLysGluThrGlyGluGluLeuGlyValLysAsnTyrGly 462  
 QY 1387 AATTTCTCTTCACCTCGCTCTATGAGCGCCATGAAGTCCATACCTACCTACCCCATC 1446  
 Db 463 ProPheThrPheLeuArgIleTyrAspAlaGlyHisMetValProTyrAspGlnProGlu 482  
 QY 1447 GCCTCCCTGCAATTTGTTTAAACGGGACTATCTTCGGT 1482  
 Db 483 AlaSerLeuGluMetValAsnSerTrpIleSerGly 494  
 RESULT 4  
 ID CBPY CANAL STANDARD; PRT; 542 AA.  
 AC P30574;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Carboxypeptidase Y precursor (EC 3.4.16.5) (Carboxypeptidase YSCY).  
 GN CPV1.  
 OS Candida albicans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93051356; PubMed=1427093;  
 RA Mukhtar M., Logan D.A., Kafer N.F.;  
 RT "The carboxypeptidase Y-encoding gene from Candida albicans and its  
 RL transcription during yeast-to-hyphae conversion.";  
 RL Gene 121:173-177(1992).  
 CC -1- FUNCTION: INVOLVED IN DEGRADATION OF SMALL PEPTIDES.  
 CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a  
 CC broad specificity.  
 CC -1- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.  
 CC -1- INDUCTION: TRANSIENTLY DOWN-REGULATED DURING THE EARLY EVENTS OF  
 CC YEAST TO HYPAE CONVERSION.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M95182; AAA34326.2; -  
 CC HSSP; P00729; 1CPY. -  
 CC MEROPS; S10.001; -  
 CC InterPro; IPR000379; Ser\_estrs\_site.  
 CC InterPro; IPR001563; Serine\_carbpept.  
 CC Pfam; PF00450; serine\_carbpept; 1.  
 CC PRINTS; PR00724; CRBOXYPTASEC.  
 CC ProDom; PD001189; Serine\_carbpept; 1.  
 CC PROSITE; PS00131; CARBOXYPEPT\_SER\_SER; 1.  
 CC PROSITE; PS00560; CARBOXYPEPT\_SER\_HIS; 1.  
 CC Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.  
 CC SIGNAL 1 21  
 FT PROPEP 22 127  
 FT CHAIN 128 542  
 FT ACT\_SITE 269 269  
 FT ACT\_SITE 461 461  
 FT ACT\_SITE 518 518  
 FT BINDING 464 464  
 FT BINDING 519 519  
 FT DISULFID 182 421  
 FT DISULFID 316 330  
 FT DISULFID 340 363  
 FT DISULFID 347 356  
 FT DISULFID 385 391  
 FT CARBOHYD 213 213  
 FT CARBOHYD 291 291  
 FT SIGNAL 542 AA; 61044 MW; 7FA6B9F82F9D44F CRC64;  
 SQ SEQUENCE

Alignment Scores:			
Pred. No.:	3,97e-34	Length:	542
Score:	599.00	Matches:	169
Percent Similarity:	43.47%	Conservative:	64
Best Local Similarity:	31.53%	Mismatches:	193
Query Match:	20.63%	Indels:	110
DB:	1	Gaps:	20
US-09-712-338-1_COPY_55_1662 (1-1608) x CBPY_CANAL (1-542)			
QY	1	CTTCCAGAGTACACCG-----GCGTCGCTCGGTAGAGACAGCTACCCCAAGAACCC	54
DB	71	LeuAspGlyLeuThrProGluLeuLysAsnIleTrpSerGluMetLeuMetLysPhePro	90
QY	55	ACCGGGGTCAAGACTCTTACAAACCCCAACAATGTCACC---ATCCGGTACAGNAACC	111
DB	91	-----AsnSerIleThrGluLeuAsnPheLysAlaPro	101
QY	112	GGGGCAGAGGGCGCTCGGAGACT-----	135
DB	102	ProLysLysGlyLysIleThrThrGlnGlnPheAspPheHisValThrAspAlaGlnVal	121
QY	136	-----ACCCCG-----GGTGCAAA	150
DB	122	ProAsnHisLysLeuArgIleLysSerThrProLysAspLeuGlyIleAspThrValLys	141
QY	151	TCCTACTCTGGATATGTCAGACCTCTCCCGAG---TCCATACCTCTCTGTTCTTC	207
DB	142	GlnTyrSerGlyTyrLeuAspValValAspGluAspLysPheThrTyrPhePhe	161
QY	208	GAAGCAGACATACCCAGAACTGCACCTATCACATGTGGTGAATGTGGCCCTCGGA	267
DB	162	GluSerArgAsnAspProLysAsnAspProValIleLeuIrpLeuAsnGlyGlyProGly	181
QY	268	AGCGATCTTTGATCGCTCTCTCGAAGAGTGTGGCCCTTGCATCATCAATCGACTTT	327
DB	182	CysSerSerLeuThrGlyLeuPhePheGluLeuGlyProSerSerIleAspLysAsnLeu	201
QY	328	GATGACTATACATCAACCTCACTCGTGAAGAGGTCTCCAAATTTACTATTCCTGCCAG	387
DB	202	LysProValTyrAsnProHisSerTrpAsnAlaAsnAlaSerValIlePheLeuAspGln	221
QY	388	CCATTGGAGTGGCTTTTCATATAGTGATACGGTTGATGGTCCATTAACCCCTGAAC	447
DB	222	ProIleAsnValGlyTyrSerTyrSer-----	230
QY	448	GGGGTCGTCGAAAATTCGAGCTTTTCAGAGAGTTACGGCCGGTACCCCAACCATGTGCC	507
DB	231	-----Ser	231
QY	508	ACTCTGATGATACTACCAATCTGCGGCAGAGCCCTGCGGAGATCCTGCAAGGATTC	567
DB	232	GlnSerValSerAsnThrIleAlaAlaGlyLysAspValTyrAlaPheLeuGlnLeuPhe	251
QY	568	CTTAGTGACTACTAGCTTGGACTCTAGGTGCAGTCTAAGGACTTCAGTCTATGGACG	627
DB	252	PheLysAsnPhePro-----GluTyrAlaAsnLeuAspPheHisIleAlaGly	267
QY	628	GAGAGCTATGAGGGGACTATGTCCTGCATCTTCAATCAATTTTACGAGCAGAAATGAG	687
DB	268	GluSerTyrAlaGlyHisTyrIleProAlaPheAlaSerGluIleLeuThrHisProGlu	287
QY	688	AGAATTCACACGGTAGTGTAAAGGTGTTTACGCTTAATTTCACTCTCTGGGAATTA	747
DB	288	Arg-----AsnPheAsnLeuThrSerValLeu	296
QY	748	-----AACGCCATCATCGAGCGGATCCAGGCCCTTACTACCTCGAATTCGCTGTG	801
DB	297	IleGlyAsnGlyLeuThrAspProLeuValGlnTyrGlnTyrTyrGluProMetAlaCys	316
QY	802	AACAATACCTACGGTATCAAGCTGTCAACGAGACCGCTCTACACTACATGAAGTTGCC	861







```

FT    HELIX      195      198
FT    STRAND     200      202
FT    TURN       208      209
FT    TURN       211      212
FT    STRAND     214      215
FT    HELIX      224      240
FT    TURN       242      243
FT    TURN       245      248
FT    STRAND     251      256
FT    TURN       257      258
FT    HELIX      259      270
FT    TURN       271      272
FT    STRAND     282      286
FT    HELIX      292      295
FT    TURN       296      298
FT    TURN       303      305
FT    HELIX      315      338
FT    HELIX      341      355
FT    TURN       356      362
FT    STRAND     365      367
FT    TURN       366      367
FT    STRAND     368      368
FT    TURN       381      381
FT    HELIX      382      391
FT    HELIX      393      398
FT    TURN       399      400
FT    STRAND     409      409
FT    HELIX      411      418
FT    TURN       419      421
FT    HELIX      422      424
FT    TURN       427      427
FT    HELIX      428      436
FT    TURN       437      438
FT    STRAND     441      446
FT    TURN       447      448
FT    TURN       451      452
FT    HELIX      454      463
FT    TURN       467      468
FT    HELIX      469      474
FT    STRAND     478      481
FT    TURN       483      485
FT    STRAND     488      492
FT    STRAND     494      495
FT    TURN       496      497
FT    STRAND     498      503
FT    TURN       504      505
FT    HELIX      510      513
FT    HELIX      515      526
FT    TURN       527      529
SQ    SEQUENCE   532 AA; 59802 MW; 7227F3489CBDD952 CRC64;

Alignment Scores:
Pred. No.:      3,79e-29      Length:      532
Score:          527.50      Matches:      149
Percent Similarity: 42.97%      Conservative: 71
Best Local Similarity: 29.10%      Mismatches: 195
Query Match:    18.16%      Indels:      97
DB:             1          Gaps:        19

US-09-712-338-1_COPY_55_1662 (1-1608) x CBPY_YEAST (1-532)

QY    43  CCCAAGACCCACCGGGGTCAGACT-----CITACACCGCAACAAT 87
Db    82  ProlysPheProGluAlaIleLysThrLysLysAspTrpAspPheValValLysAsnAsp 101
QY    88  GTCACCATCCGGTAC-----AAGGAACCCGGGCGACAGGGGTC 126
Dd    102  AlaIleGluAsnTyrGlnLeuArgValAsnLysIleLysAspProLysIleLeuGlyLeu 121
QY    127  TGCAGACTACCCCGGGTGTCAAAATCCTACTCTGGATATGTCGACACCTCTCCCGAG--- 183

```

```

Dd    122  -----AspProAsnValThrGlnTyrThrGlyTyrLeuAspValGluAspGluAsp 138
QY    184  TCCCAIACCTTCCTTCCTTCGAGCCAGACATAACCCAGAAACGCACTTACACA 243
Dd    139  LysHisPhePhePheTyrThrPheGluSerArgAsnAspProAlaLysAspProValIle 158
QY    244  TGTGGTTGAATGGTGGCCCTGGAGAGCATTCCTTGTATCGGTCCTCTCGAAGATTGGGC 303
Dd    159  LeuTrpLeuAsnGlyGlyProGlyCysSerSerLeuThrGlyLeuPhePheGluLeuGly 178
QY    304  CCTTGCATGTCAATTCGACTTTTGATGACTACATCAACCTCACTCGTGGCAACAGGTC 363
Dd    179  ProSerSerIleGlyProAspLeuLysProIleGlyAsnProTyrSerTyrAsnSerAsn 198
QY    364  TCCAATTTACTATTCCTTCCTCCAGCCATTCGGAGTCGGCTTTTCATATGATGATACGGTT 423
Dd    199  AlaThrValIlePheLeuAspGlnProValAsnValGlyPheSerIyrSer----- 215
QY    424  GATGGGTCCATTAACTTAACTGGGGTGGTGGGAAATTCGAGCTTTGAGGAGTTCAG 483
Dd    216  -----GlySerSer 218
QY    484  GGCCGGTACCCAAACCAATTGATGCCACTCTGATCGATACTACCAATCTTGGCGAGAGGCC 543
Dd    219  Gly-----ValSerAsnThrValAlaIleGlyLysAsp 229
QY    544  GCTTGGAGATCCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGGTGCAG 603
Dd    230  ValTyrAsnPheLeuGluLeuPhePheAspGlnPheProGluTyrValAsnLysGlyGln 249
QY    604  TCTAAGGACTTCAGTCTATGGCGGAGAGCTATGAGGGGCACATGCTGCTGATTCCTTC 663
Dd    250  -----AspPheHisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheAla 267
QY    664  AATCATTTTTACGAGCAGAAATGAGACAATGCCAACGGTAGTGTAAIGTGTTCAGCTT 723
Dd    268  SerGluIleLeuSerHisLysAspArg----- 276
QY    724  AATTTCAACTCTCTGGGAATTAIT-----AAGGCATCATGAGCAGGCGATCCAGGCC 777
Dd    277  AsnPheAsnLeuThrSerValLeuIleGlyAsnGlyLeuThrAspProLeuThrGlnTyr 296
QY    778  CATTACTACCTCAATTCGCT-----GTGAACAATACCTAC 813
Dd    297  AsnTyrTyrGluProMetAlaCysGlyGlyGluGlyGluProSerValLeuProSerGlu 316
QY    814  GGTATCAAGGCTGTCAACGAGACCGTCTACACTCATCAAGTTGCCAACCAATGCCA 873
Dd    317  GluCysSerAlaMetGluAspSerLeu----- 325
QY    874  AATGTTGCCAGGATTTGATTTCCACCTGCAACACAGACAAACCGCAGCATTAGCTGAC 933
Dd    326  GluArgCysLeuGlyLeuIleGluSerCysTyrAspSerGln-----Ser 340
QY    934  TACGCCCTCTGGCCGAGACCCACCAACATGTCAGGAGCAATGTTGAGGGCCATCTAC 993
Dd    341  ValTrpSerCysValProAlaThrIleTyrCysAsnAsnAlaGlnLeuAlaProTyrGln 360
QY    994  GCCTTTGCTGGTGGTGTGTATGATATTCGGCATCCATATGATGAC----- 1041
Dd    361  Arg---ThrGlyArgAsnValTyrAspIleArgLysAspCysGluGlyAsnLeuCys 379
QY    1042  ---CCGACTCCGCCAAGTTATTACAAACAATTTCTGGCAAGGACTCTGTCTAGGACGCT 1098
Dd    380  TyrProThrLeuGlnAsp---IleAspAspTyrLeuAsnGlnAspTyrValLysGluAla 398
QY    1099  ATCCGGCTCACATCACTACACCCAGTCC---AATAATGACGCTCTACTACGCTTTCCAG 1155
Dd    399  ValGlyAlaGluValAspHisTyrGluSerCysAsnPheAspIleAsnArgAsnPheLeu 418
QY    1156  CAAACAGCGGACTTGTGTCGCCCACTTCATCCAGACCTTCGAGGAGATCCTTCGCTCTC 1215
Dd    419  PheAlaGlyAspTrpMetLysPro---TyrHisThrAlaValThrAspLeuLeuAsnGln 437

```

Nature 408:820-822(2000).  
[3]

Db 172 AsnLeuLeuTyrValAspGlnProValGlyThrGlyPheSerTyrThrAspLysSer 191  
 QY 427 GGGTCCATTAAACCTGTACTGGGTGCGICGAAATTCGAGCTTTCAGGAGTTCAGGC 486  
 Db 192 AspIleArgHisAspGluThrGlyVal 200  
 QY 487 CGGTACCCACCATGTGCGCACTCTGATCGATACCTACCAATCTTGCCGAGAGCGGCT 546  
 Db 201 200  
 QY 547 TGGGAGATCCTGCAAGGATTCCTTAGTGACTACCTAGCTTGGACTTACGGTGCAGTCT 606  
 Db 205 TyrAspPheLeuGlnAlaPheAlaGluHisProLysLeu 220  
 QY 607 AAGGACTTCTAGTCTATGAGCGAGAGCTATGAGGCGCACTATGCTGCTGATCTTCAAT 666  
 Db 221 AsnAspPheTyrLileThrGlyGluSerTyrAlaGlyHisTyrLileProAlaPheAlaSer 240  
 QY 667 CATTTTTACGACGATGAGAGAATTCGCAACGGTAGTCTTAATGGTGTTCAGCTTAAT 726  
 Db 241 ArgValHisTysGlyAsnLys 255  
 QY 727 TTCACCTCTCTGGATTTTAAACGGCATAACGAGCGGATGAGCGGCGCTTACTAC 786  
 Db 256 LeuLysGlyPheAlaIleGlyAsnGlyLeuThrAspProAlaLeuGlnTyrProAlaTyr 275  
 QY 787 CTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGCGGCTACAC 846  
 Db 276 ProAspTyrAlaLeuGlu 289  
 QY 847 IACATCAAGTTTCCCAACCAATGCAATGGTTCGAGGATTTGATTTCCACTGCAAA 906  
 Db 290 HisAspArgLeuGlnLysIleValProLeu--CysGluLeuSerIleLysLeuCysGly 308  
 QY 907 CAGCAAAACCGCAGCGCA--TTAGCTGACTAGCCCTCTGCGCGCAACCCACACACG 963  
 Db 309 ThrAspGlyThrThrSerCysLeuAlaSerTyrLeuValCysAsnSerLeuPheSerGly 328  
 QY 964 TCCAGGACAATGTTGAGGGCCATACCTACGCGCTTTCGCTGCTGTGTATGATAT 1023  
 Db 329 ValMetSerHisAlaGlyGlyValAsnTyr 341  
 QY 1024 CGGCATCCA 357  
 Db 342 ArgLysLysCysValGlySerLeuCysTyrAspPhe 377  
 QY 1066 AAATTTCTGGCAAGACTCTGTCTATGAGCGTATCGCGTC--AACATCAACATACACC 1122  
 Db 358 LysPheLeuAsnLeuGlnSerValArgLysSerLeuGlyValGlyAspIleAspPheVal 377  
 QY 1123 CAGTCCATATAGCGTCTACTAGCTTTCCAGCAACAGCGGACTTGTCTGCGCCCAAC 1182  
 Db 378 SerCysSerThrSerValTyrGlnAla 1230  
 QY 1183 TTCATGAGACCTCGAGGAGATCTTCTCTCCCGTGGTGTCTCC--MetLeuValAsp 390  
 Db 391 TrpMetArgAsnLeuGlu 408  
 QY 1231 ---CTCATCTAT---GGCAGCGCGCATACATCTGCACTGCTTCCGCGGTCAGCCGCTT 1284  
 Db 409 LeuLeuValIleAlaGlyGluTyrAspLeuIleCysAsnTrpLeuGlyAsnSerArgTrp 428  
 QY 1285 TCCCTCGTGGCACTACTCCCAAGCGCCCGAGTTCGAGCGGAGGAGTACACGCCGCG 1344  
 Db 429 ValAsnAlaMetGluTyrSerGlyLysThrAsnPheGlyAlaAlaLysGluValProPhe 448  
 QY 1345 AAGTCAACGCGTCGATGCGGAACTCGGAGTATGTAATTTCTCTCTCACTGCG 1404  
 Db 449 IleValAspGlyLysGluAlaGlyLeuLeuLysThrTyrGluGlnLeuSerPheLeuLys 468  
 QY 1405 GTCTATGAGCGGCGGCAAGTGAAGTCCCATCTACAGCGCCATCGCTCCCTCCCAATTT 1464  
 Db 469 ValArgAspAlaGlyHisMetValProMetAspGlnProLysAlaAlaLeuLysMetLeu 488

QY 1465 AACCGG 1470  
 Db 489 LysArg 490  
 RESULT 8  
 ID CBPX\_ORYSA STANDARD; PRT; 429 AA.  
 AC PS2712;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Serine carboxypeptidase-like precursor (EC 3.4.16.-).  
 GN CBP31.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoideae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Yukihikari;  
 RA Washio K., Ishikawa K.;  
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a  
 broad specificity.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D17587; BAA04511.1; -;  
 CC PIR: T03607; T03607.  
 CC HSP: P00729; IYSC.  
 CC MEROPS: S10.009; -;  
 CC Gramene: P52712; -;  
 CC InterPro: IPR000379; Ser\_estr\_site.  
 CC InterPro: IPR001563; Serine\_carpept.  
 CC Pfam: PF00450; serine\_carpept; 1.  
 CC PRINTS: PR00724; CRBOXYPTASEC.  
 CC ProDom: PD001189; Serine\_carpept; 1.  
 CC PROSITE: PS00131; CARBOXYPEPT\_SER\_SER; 1.  
 CC PROSITE: PS00560; CARBOXYPEPT\_SER\_HIS; 1.  
 KW Hydrolase; Carboxypeptidase; Glycoprotein; Signal.  
 FT SIGNAL 1 ? POTENTIAL.  
 FT CHAIN ? 429 SERINE CARBOXYPEPTIDASE-LIKE.  
 FT ACT\_SITE 148 148 BY SIMILARITY.  
 FT ACT\_SITE 336 336 BY SIMILARITY.  
 FT ACT\_SITE 393 393 BY SIMILARITY.  
 FT BINDING 339 339 SUBSTRATE (BY SIMILARITY).  
 FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 429 AA; 47746 MW; 1D5A668544325BB1 CRC64;  
 Alignment Scores:  
 Pred. No.: 2,23e-25 Length: 429  
 Score: 473.00 Matches: 130  
 Percent Similarity: 45.49% Conservative: 77  
 Best Local Similarity: 28.57% Mismatches: 178  
 Query Match: 16.29% Indels: 70  
 DB: 1 Gaps: 16  
 US-09-712-338-1\_COPY\_55\_1662 (1-1608) x CBPX\_ORYSA (1-429)

QY 154 TACTCTGGAATAT-----GTGACACCTCTCCCGAGTCCCATCTTCTTCTGTTCTTC 207  
 Db 17 HisalaglyTyrTyrArgLeuProAsnThrHisAspAlaArgLeuPheTyrPhePhePhe 36

```

QY 208 GAAGCCAGACATACCA---GAACTGCACTATACATTTGTTGTTGAATGGGGCCCT 264
Db 37 GluSerArgGlySerLysGlyGluAspProValValIleThrLeuThrGlyGlyPro 56
QY 265 GGAAGCGATTCTTTGATCGGCTCTTGAAGAGTGGGCCCTGCCAFGTCAATTCGACT 324
Db 57 GlyCysSerSerGluLeuAlaLeuPheThrGluAsnGlyProPheHisIleAlaAspAsn 76
QY 325 TTTCATGACTACATACCCCTACTCTGGGAACGAGGTCTCAATTTACTATTCCTGTC 384
Db 77 MetSerLeuValTyrAsnAspPheGlyTyrAspGlnGluSerAsnLeuIleThrValAsp 96
QY 385 CAGCCATTGGAGTCGGCTTTTCATATAGTATAGGTGATGGTGCATTAACCCCTGTA 444
Db 97 GlnProThrGlyThrGlyPheSerTyrSerAsnProArgAspIleThrArgHisAspGlu 116
QY 445 ACTGGGTGCGTCGAAATTCAGAGCTTTCAGAGAGTTCAGGCGCGGTACCAACCAATGAT 504
Db 117 AlaGly---ValSerAsnAspLeuTyrAla----- 125
QY 505 GCACTCTGATCGATACATACCAATCTTCCCGCAGAGCGCGCTTGGAGATCCTGCAAGA 564
Db 126 -----PheLeuGlnAla 129
QY 565 TTCCTTAGTGGACTACCTAGCTTGGACTCTAGGTGTCAGTCTAAGGACTTCAGTCTATGG 624
Db 130 PhePheThrGluHisProAsnPhe-----AlaLysAsnAspPheTyrIleThr 145
QY 625 ACGGAGAGCTATGGGCGACTATGCTCTGCTGATCTTCAATTCATTTTACGAGCAGAT 684
Db 146 GlyGluSerTyrAlaGlyHisTyrIleProAlaPheAlaSerArgValTyrLysGlyAsn 165
QY 685 GAGCAATTTGCCAAGCTAGTGTATGATGTGTTCACCTTAATTCACCTCTCTGGAAAT 744
Db 166 Lys-----AsnSerGluGlyIleHisIleAsnLeuLysGlyPheAlaIle 180
QY 745 ATTACGGCATCATCGACGAGCGATCCAGGCGCCCTTACTACCTCAATTCGCTGTGAAC 804
Db 181 GlyAsnGlyLeuThrAspProAlaIleGlnTyrLysAlaTyrThrAspTyrSerLeuAsp 200
QY 805 AATACCTAGGTATCAAGCTGTCAACGAGACCGCTCTACAACTACATGAGATTTGCCAAC 864
Db 201 -----MetGlyLeuIleThrLysSerGlnPheAsnArgile-----Asn 213
QY 865 CAATGCCAAATGTTGCCAGGATTTGATTTCCACCTGCAACAGACAAACCGC---ACC 921
Db 214 LysIleValProThrCysGluLeuAlaIleLysLeuCysGlyThrSerGlyThrIleSer 981
QY 922 GCATTAGCTGACTACGCCCTCTCGCGGAGACCAACCAATGTCAGGGAATGTGAG 981
Db 234 CysLeuGlyAlaTyrValValCys-----AsnLeuIleCysSerSerIleGlu 249
QY 982 GGGCCATACACGCTTTGCTGCTGCTGCTATGATCAATTCGCAATCA----- 1032
Db 250 -----ThrIleGlyLysLysAsnTyrTyrAspIleArgLysProCysValGly 266
QY 1033 -----TATGATGACCGACTCCGCCAAGTATTACAAATTTCTGGCAAGGAC 1083
Db 267 SerLeuCysTyrAsp-----LeuSerAsnMetGluLysPheLeuGlnLeuLys 282
QY 1084 TCTGTCATGACGCTATCGGCGTC---AACATCAACTACACCCAGTCCCAATATGACGTC 1140
Db 283 SerValArgGluSerLeuGlyValGlyAspIleGlnPheValSerCysSerProThrVal 302
QY 1141 TACTACGCTTCCAGCAACAGCGGACTTTGTCGCGCCCACTTCATCGAA---GACCTC 1197
Db 303 TyrGlnAlaMetLeu-----LeuAspTrpMetArgAsnLeuGluValGlyIle 318
QY 1198 GAGGAGATCTTGTCTCTCCCGTCTCCCTCATCTATGCGGAGCGGATATGATC 1257
Db 319 ProGluLeuLeuGluAsnAspIleLysValIleThrAlaGlyGluTyrAspLeuIle 338

```

---

```

QY 1258 TGCRACTGTTGCGGGTTCAGCCGCTTTCCTCGCTGCGAACTACTCCCAAGCCGCCAG 1317
Db 339 CysAsnTrpLeuGlyAsnSerArgTrpValAsnSerMetGluTrpSerGlyLysGluAla 358
QY 1318 TTCGAAGCGAGGTACACGCCCTGAAAGTCAACGGCTGAGTATGGGAAACTGCG 1377
Db 359 PheValSerSerSerGluGluProPheThrValAspGlyLysGluAlaGlyIleLeuLys 378
QY 1378 GAGTATGCTAATTTCTCTTCTACTCGCTATGAGGAGCGGAGTCCCAAGTCCCATAC 1437
Db 379 SerTyrGlyProLeuSerPheLeuLysValHisAspAlaGlyHisMetValProMetAsp 398
QY 1438 CAGCCCATCGCTCTCCCTGCAATTTTACCGGACTATCTTCGCT 1482
Db 399 GlnProLysValAlaLeuGluMetLeuMetArgTrpThrSerGly 413

```

RESULT 9

CBP3\_ORYSA STANDARD; PRT; 500 AA.

ID CBP3\_ORYSA

AC P37891;

DT 01-OCT-1994 (Rel. 30, Created)

DI 01-OCT-1994 (Rel. 30, Last sequence update)

DI 28-FEB-2003 (Rel. 41, Last annotation update)

DE Serine carboxypeptidase III precursor (EC 3.4.16.5).

GN CBP3.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzaceae; Oryza.

OX NCBI\_TaxID=4530;

RN [1]

RC SEQUENCE FROM N.A.

RC STRAIN=cv. Yukihikari; TISSUE=Seed;

RC MEDLINE=92329723; PubMed=1627776;

RA Washio K., Ishikawa K.;

RT "Structure and expression during the germination of rice seeds of the

RT gene for a carboxypeptidase";

RL Plant Mol. Biol. 19:631-640(1992).

CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a

CC broad specificity.

CC -1- SUBUNIT: Monomer (Probable).

CC -1- INDUCTION: BY GIBBERELIC ACID (GA). INHIBITED BY ABSICISIC ACID

CC (ABA).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; D10985; BAA01757.1; -

DR PIR; S22530; S22530.

DR HSP; P00729; IYSC.

DR MEROPS; S10.009; -.

DR Gramene; P37891; -.

DR InterPro; IPR000379; Ser\_estrs\_site.

DR InterPro; IPR001563; Serine\_carbpept.

DR Pfam; PF00450; serine\_carbpept; 1.

DR PRINTS; PR00724; CRBOXPTASEC.

DR PRODOM; PD001189; Serine\_carbpept; 1.

DR PROSITE; PS00131; CARBOXYPEPT\_SER\_SER; 1.

DR PROSITE; PS00560; CARBOXYPEPT\_SER\_HIS; 1.

DR Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.

KW SIGNAL

FT SIGNAL 1 21

FT PROPEP 22 73 BY SIMILARITY.

FT CHAIN 74 484 SERINE CARBOXYPEPTIDASE III.

FT PROPEP 485 500 BY SIMILARITY.

FT ACT\_SITE 216 216 BY SIMILARITY.

FT ACT\_SITE 404 404 BY SIMILARITY.

FT ACT\_SITE 461 461 BY SIMILARITY.

FT	BINDING	407	SUBSTRATE (BY SIMILARITY).
FT	CARBOHYD	144	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	500 AA; 55446 MW; AE455E2780147D88 CRC64;	
Alignment Scores:			
Pred. No.:	6,79e-24	Length:	500
Percent:	452.00	Matches:	133
Best local Similarity:	43.56%	Conservative:	80
Query Match:	27.20%	Mismatches:	206
DB:	15.56%	Indels:	70
	1	Gaps:	17
US-09-712-338-l_COPY_55_1662 (1-1608) x CBP3_ORYSA (1-500)			
QY	40	CTACCCAAAG-----AACCCCAACCGGG-----	GTCAAGACTCTTTACACCGCAAAAC 84
Db	47	LeuProLysGluAlaGlyProThrGlyAlaGlyAspValProSerValAlaProGlyGlu 66	
QY	85	AATGTCACCAACCGGTACAGAAACCCGGG---GCAGAGGGGTCTGCGAGACTACCCCG 141	
Db	67	LeuLeuGluArgValThrLeuProGlyLeuProGlyValclyAsp----- 83	
QY	142	GGTGTCAAATCCIACTCTGGATAT-----GTCGACACCTCTCCGAGTCCCATACCTTC 195	
Db	84	--LeuGlyHisHisAlaGlyTyrTyrArgLeuProAsnThrHisAspAlaArgMetPhe 102	
QY	196	TTCTGGTCTTCGAAGCCAGACATACCCAGAAACTGCACCTATCACATTGGTGGTGAAT 255	
Db	103	TyrPheLeuPheGluSerArgGlyLysLysGluAsp---ProValValIleIlePleuThr 121	
QY	256	GGTGGCCCTCGAACGGATCTTTCGATCGTCTCTTCGAAGAGTTGGGCCCTTCGCATGTC 315	
Db	122	GlyGlyProGlyCysSerSerGluLeuAlaValPheTyrGluAsnGlyProPheThrIle 141	
QY	316	AATTGCTACTTTTGATGACTACATCAACCCCTCCTCGTGGACGAGGTCTCCAAATTACTA 375	
Db	142	SerAsnAsnMetSerLeuAlaTrpAsnLysPheGlyTrpAspThrIleSerAsnIleIle 161	
QY	376	TTCTGTCCCGACCATTTGGGAGTCGGCTTTTCATATAGTGATACGTTGATGGGTCCATT 435	
Db	162	PheValAspGlnProThrGlyThrGlyPheSerTyrSerSerAspAspArgAspThrArg 181	
QY	436	AACCCCTGAACCTGGGTGTCGCGAAATTCGAGCTTTCGAGAGTTTCAGGGCGGTACCCA 495	
Db	182	HisAspGluThrGlyVal----- 187	
QY	496	ACCATTGATGCCACTCTGATCGATACCTACCAATCTTCCGCGACAGGCGCTTGGAGATC 555	
Db	188	-----SerAsnAspLeuTyrSerPhe 194	
QY	556	CTCGAGGATTCCTTAGTGGACACCTAGCTTGGACTCIAGGGTGCAGTCTAAGGACTTC 615	
Db	195	LeuGluValPhePheLysLysHisProGluPhe-----AlaLysAsnAspPhe 210	
QY	616	AGTCTATGGACGGAGACTATGGAGGCACATATGSICTGCAITCTTCAATCAATTTTAC 675	
Db	211	PheIleThrGlyGluSerTyrAlaGlyHisTyrIleProAlaPheAlaSerArgValHis 230	
QY	676	GACGAGAATTCAGAGAATTGCCACGGTAGTGTTAATGGTGTACGCTTAATTTCAACTCT 735	
Db	231	GinglyAsnLys-----AlaAsn-----GluGlyIleHisIleAsnLeuLysGly 245	
QY	736	CTVGGGAATATTAAACGGCATTCAGACGGCGATCCAGGCCCTTACTACCTCGAATTC 795	
Db	246	PheAlaIleGlyAsnGlyLeuThrAspProAlaIleGlnTyrLysAlaIleThrAspTyr 265	
QY	796	GCTGTCAACAATACCTACGGTATCAAGCTGTCAACGACCGTCTACACTACATGAAG 855	
Db	266	AlaLeuAsp-----MetAsnLeuIleLysLysSerAspTyrAspArgIle--- 280	
QY	856	TTTGCCAACCAATGCCAAATGTTGCCAGGATTTTGATTTCCACCTTCGAACAGACAAC 915	
Db	281	-----AsnLysPheIleProCysGluPheAlaIleLysLeuCysGlyThrAsnGly 298	





```

RESULT 11
CBP3_WHEAT
ID CBP3_WHEAT STANDARD; PRT; 500 AA.
AC P11515;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y.";
RL J. Biol. Chem. 262:13726-13735(1987).
CC -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -!- SUBUNIT: Monomer (Probable).
CC -!- INDUCTION: By gibberellic acid (GA).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; J02817; AAA34273.1; -
DR PIR; A29412; A29412.
DR HSP; P00729; LCPY.
DR MEROPS; S10.009; -.
DR InterPro; IPR000379; Ser_estrs_site.
DR InterPro; IPR001563; Serine_carbpept.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXPTASEC.
DR ProDom; PD001189; Serine_carbpept; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 21
FT PROPEP 22 73 BY SIMILARITY.
FT CHAIN 74 484 SERINE CARBOXYPEPTIDASE III.
FT PROPEP 485 500 BY SIMILARITY.
FT ACT_SITE 216 216 BY SIMILARITY.
FT ACT_SITE 404 404 BY SIMILARITY.
FT ACT_SITE 461 461 BY SIMILARITY.
FT BINDING 407 407 SUBSTRATE (BY SIMILARITY).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 500 AA; 55334 MW; B2ACE10EF9484CDA CRC64;
SQ

```

```

Alignment Scores:
Pred. No.:      3 27e-21      Length:      500
Score:          413.50      Matches:     136
Percent Similarity: 41.37%      Conservative:  80
Best Local Similarity: 25.30%      Mismatches:   217
Query Match:      14.24%      Indels:       75
DB:               1          Gaps:          16

US-09-712-338-1_COPY_55_1662 (1-1608) x CBP3_WHEAT (1-500)

Qy      1  CTTCCAGGAAGTACCCGCGTCCGTCGTAGACACAGCTACCAAGAACCCACCGGG 60
      |||||
Db      47  LeuProGlyArgProArgGlyLeuGly-----AlaGly 58

```

61 GTCACAGACTTTCACACCGCAAAATGTCACCATCCGGTAGCAAGAACCCGGG---GCA 117  
Db : : : : :  
59 AlaGluAspValAlaProGlyGlnLeuLeuGluArgValThrLeuProGlyLeuPro 78  
QY : : : : :  
118 GAGGGCGTGTCGCAGACTACCCGGGTGTCAAAATCCTACTCTGGATAT-----GTCGAC 171  
Db ||||| : : : : :  
79 GluGlyValGlyLysP-----LeuGlyHisHisAlaGlyTyrItyrArgLeuPro 94  
QY : : : : :  
172 ACCTGCCGAGTCCCATAACCTTCTGCTGTTCTCGAAGCCGACGACAATAACCCAGAAACT 231  
Db : : : : :  
95 AsnThr-HisAspAlaArgMetPheTyrPhePhePheGluSerArgGlyLysIysGluAsp 114  
QY : : : : :  
232 GCACCTATCACATTGNGTGTGAATGGTGCCTCGAAGCGATTCTTTGATCGTCTCTTC 291  
Db : : : : :  
115 ---ProValValIlePrLeuthrGlyProGlyCysSerSergluLeuAlaValPhe 133  
QY : : : : :  
292 GAAGAGTTCGGCCCTGCCATGCAATTCGACTTTTGATGACTACATCAACCTCACATCG 351  
Db ||||| : : : : :  
134 TyrGluAsnGlyProPheThrIleAlaAsnAsnMetSerLeuValTrpAsnIysPheGly 153  
QY : : : : :  
352 TGAACGAGAGTCTCCAATTACTATTCCTCCAGCCATCGGAGTGGCGTTCATCAT 411  
Db : : : : :  
154 TrpAspLysIleSerAsnIleIlePheValaspProAlaThrGlyThrGlyPheSerItyr 173  
QY : : : : :  
412 AGTGATACGGTGTGNGGTCCTAATTAACCCIGTPAACITGGGTGCTGAAAATTCGAGCTTT 471  
Db ||||| : : : : :  
174 SerSeraspArgAspThrArghisAspGluAlaGlyVal----- 187  
QY : : : : :  
472 GCAGGAGTTCAGGGCGGTACCCACCAATGATGCCACTCTGATCGCATACTACCAATCTT 531  
Db ----- 187  
532 GCCGAGAGCGCGTGTGGAGATCCTGCAAGGATTCTTAGTGGACTACCTAGCTTGGAC 591  
QY : : : : :  
188 ---SerAsnAspLeuTyraSpPheLeuGlnValPhePheLysLysHisProGluPhe--- 205  
QY : : : : :  
592 TCTAGGGTCAGTCTAAGGACTTCAGTCTATGGACGGAGACTATGGAGGCACATATGGT 651  
Db -----ValLysAsnAspPhePheIlethrGlyGluSerTyralaglyHistyrlle 222  
652 CDTGCATTCTCAATCATTTTTACGAGCAGAATGAGAAATGCCACGGTAGTGTAAAT 711  
QY ||||| : : : : :  
223 ProAlaPheAlaSerArgValHisGInglyAsnLysLys-----AsnGlu 237  
QY : : : : :  
712 GGTTGTCAGCTTAATTTCAACTCTCTGGGAATATTAAACGGCATCATCGAGAGCGCATC 771  
Db ||||| : : : : :  
238 GlyThrHisIleAsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAspProAlaIle 257  
QY : : : : :  
772 CAGCCCCCTTACTACCTCGUANTCCGTCGTGAACAAATACCTACGGTATCAAGGTGTCAAC 831  
Db GlnTyrLysAlaTyrThrAspTyralalaLeuAsp-----MetAsnLeuIleGln 273  
QY : : : : :  
832 GAGACCGTCTACAACACTACATGAAGTTTCCCAACCAAATGCCAAATGGTGTGCCAGGATTG 891  
Db : : : : :  
274 LysAlaasPTyrAspArgIle-----AsnLysPheIleProProCysGluPheAla 290  
QY : : : : :  
892 ATTTCCACCTCGAAACACACAAACCGC---ACCGCATTAGTCTACTACGCCCTCTCGGCC 948  
Db ||||| : : : : :  
291 IleLysLeuCysGlyThrAspGlyLysAlaSerCysMetalaAlaTyrMetValCysAsn 310  
QY : : : : :  
949 GARCCCAACCATGTCAGGGCAATGTTGAGGGGCCCATACTACGCTTGTGTGGTFCGT 1008  
Db ||||| : : : : :  
311 SerIlePheAsnSerIleMetLysLeuValGlyThrLysAsnTyr----- 325  
QY : : : : :  
1009 GGTGTGTATCATATTCGGCATTCATATGATGACCCGACTCCGGCCAAGTATTACAAC--- 1065  
Db ||||| : : : : :  
326 -----TyrAspValArgLysGluCysGlyGluGlyLysLeuCystyrAspPheSerAsnLeu 343  
QY : : : : :  
1066 ---AAATHTCTGCCAAGGACTCTGTCTATGGACGCTATCGCGCTC---AACATCAACTAC 1119  
Db ||||| : : : : :  
344 GluLysPhePheGlyAspLysAlaValArgGluAlaIleGlyValGlyAspIleGluPhe 363  
QY : : : : :  
1120 ACCAGTCCAAATAATGAGTCTACTACGCTTTCAGCAACAGGCGACTTGTCTGGCC 1179

Db 364 ValSerCysSerThrSerValtyrGlnAlaMetLeuThrAsp-----TrpMet 379  
 QY 1180 AACTTCATPCGAA---GACCTCGAGAGAGATCTGCTCTCCCGCGGTGCTCCCTCATC 1236  
 Db 380 ArgAsnLeuGluValGlyIleProAlaLeuLeuGluAspGlyIleAsnValLeuIleTyr 399  
 QY 1237 TATGGCAGCGGATTACATCTGCAACATCGTTTCGGCGGTGAGCGGTTTCCTCGCTCGC 1296  
 Db 400 AlaGlyGluTyrAspLeuLeuGlyCysAsnTrpLeuGlyAsnSerArgTrpValHisSerMet 419  
 QY 1297 AACTTACTCCAGCGCGCCAGCTTCGGAAGCGGAGGTACAGCCCTGAAAGTCAACGGC 1356  
 Db 420 GluTrpSerGlyGlnIleAspPheAlaLysThrAlaGluSerPheLeuValAspAsp 439  
 QY 1357 CTCGAGTATGGGAAACATCGCGAGTATGTAATTCCTCTCACTCGGCTCTATGAGCA 1416  
 Db 440 AlaGlnAlaGlyValLeuLysSerHisGlyAlaLeuSerPheLeuLysValHisAsnAla 459  
 QY 1417 GGCATGAGTCCCATCTACTACAGCCATCGCTCCCTGCAATTTGTTAAACGG 1470  
 Db 460 GlyHisMetValProMetAspGlnProLysAlaAlaLeuGluMetLeuArgArg 477

RESULT 12

CPVL\_HUMAN STANDARD; PRT; 476 AA.  
 ID CPVL\_HUMAN Q9H3G5; O8NBL7; Q9H3A1;  
 AC 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Probable serine carboxypeptidase CPVL precursor (EC 3.4.16.-)  
 DE (Carboxypeptidase, vitellogenic-like) (Vitellogenin-like) (Vitellogenin-like)  
 DE like protein (VCP-like protein).  
 DE CPVL OR VLP.  
 OS Homo sapiens (Human).  
 GN Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RN MEDLINE-21295045; PubMed-11401439;  
 RX Mahoney J.A., Ntoliou B., Dasilva R.P., Gordon S., McKnight A.J.;  
 RA "Cloning and characterization of CPVL, a novel serine  
 RT carboxypeptidase, from human macrophages.";  
 RL Genomics 72:243-251(2001).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP Cho J.-J., Baik H.-H.;  
 RA "Cloning of VCP-like protein expressed in human heart and placenta.";  
 RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RN SEQUENCE FROM N.A.  
 RP Tissue=Placenta;  
 RA Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,  
 RA Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,  
 RA Nagahara K., Sugano S., Isogai T.;  
 RA "HRI human cDNA sequencing project.";  
 RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RN SEQUENCE FROM N.A.  
 RP Tissue=Lung;  
 RC MEDLINE-22388257; PubMed-12477932;  
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RL human and mouse cDNA sequences.";  
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC !- FUNCTION: May be involved in the digestion of phagocytosed  
 CC particles in the lysosome, participation in an inflammatory  
 CC protease cascade, and trimming of peptides for antigen  
 CC presentation.  
 CC !- TISSUE SPECIFICITY: Expressed in macrophages but not in other  
 CC leukocytes. Abundantly expressed in heart and kidney. Also  
 CC expressed in spleen, leukocytes, and placenta.  
 CC !- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF106704; AAG37991.2; -  
 CC EMBL: AF282617; AAG14348.1; -  
 CC EMBL: AK075433; BAC11618.1; -  
 CC EMBL: BC016838; AAH16838.1; -  
 CC HSP: P10619; IIVY.  
 CC MEROPS: S10.003; -  
 CC Gene: HGNC:14399; CPVL.  
 CC InterPro: IPR000379; Ser\_estrs\_site.  
 CC InterPro: IPR001563; Serine\_carbpept.  
 CC Pfam: PF00450; serine\_carbpept; 1.  
 CC PRINIS: PR00724; CR00XPASEC.  
 CC PROSITE: PD001189; Serine\_carbpept; 1.  
 CC PROSITE: PS00131; CARBOXYPEPT\_SER\_SER; 1.  
 CC PROSITE: PS00560; CARBOXYPEPT\_SER\_HIS; FALSE\_NEG.  
 CC Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; signal.  
 CC SIGNAL 1 22 POTENTIAL.  
 CC PROPEP 23 ? POTENTIAL.  
 CC CHAIN 23 476 PROBABLE SERINE CARBOXYPEPTIDASE CPVL.  
 CC ACT\_SITE 204 204 BY SIMILARITY.  
 CC ACT\_SITE 388 388 BY SIMILARITY.  
 CC ACT\_SITE 448 448 BY SIMILARITY.  
 CC CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 346 346 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CONFLICT 25 25 R -> H (IN REF. 2).  
 CC CONFLICT 284 284 L -> F (IN REF. 3 AND 4).  
 CC CONFLICT 287 287 F -> L (IN REF. 3).  
 CC CONFLICT 398 398 H -> R (IN REF. 3 AND 4).  
 CC CONFLICT 422 422 F -> L (IN REF. 2).  
 CC CONFLICT 435 435 A -> V (IN REF. 3 AND 4).  
 CC CONFLICT 438 438 F -> S (IN REF. 2).  
 CC SEQUENCE 476 AA; 54110 MW; 2D966683A4F3FD01 CRC64;

Alignment Scores:  
 Pred. No.: 2,03e-17 Length: 476  
 Score: 359.00 Matches: 130  
 Percent Similarity: 39.51% Conservative: 62  
 Best Local Similarity: 26.75% Mismatches: 178  
 Query Match: 12.36% Indels: 116  
 DB: 1 Gaps: 21  
 US-09-712-338-1\_COPY\_55\_1662 (1-1608) x CPVL\_HUMAN (1-476)  
 QY 139 CCGGCT-----GTCAATCTCTCTCTGATAT-----GTGACACCTCTCCGATCC 186  
 Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86  
 ||||| :|||||:|||||:|||||: :||| :|||  
 ||||| :|||||:|||||:|||||: :||| :|||



DR	PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.	QY	154	TACTCTGATATGCGACACTCTCCCGAG-----TCCCATACCTCTTCTGTGTTCTTC	207
DR	PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.	QY	154	TACTCTGATATGCGACACTCTCCCGAG-----TCCCATACCTCTTCTGTGTTCTTC <td>207</td>	207
KW	Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Multigene family.	Db	16	TyrAlaGlyTyrValThrValSerGluAspArgGlyAlaAlaLeuPheTyrTrpPhe	35
FT	NON_TER 1	QY	208	GAAGCCAGACATAACCCAGAAATCGACCTATCATATGTGGTGTGAATGGTGGCCCTGGA	267
FT	CHAIN <1 256	Db	36	GluAlaAlaHisAspProAlaSerLysProLeuLeuLeuTrpLeuAsnGlyGlyProGly	55
FT	PROPEP 257 270	QY	268	AGCGATTCTTTG---ATCGGTCTCTTGAAGAGTITGGCCCTTCCCATGCAATTCGACT	324
FT	CHAIN 271 436	Db	56	CysSerSerIleAlaPheGlyValGlyGluGluValGlyProPheHisValAsnAlaAsp	75
FT	ACT_SITE 149 149	QY	325	TTTGATGAC---TACATCAACCCCTCACTCGGTGAACGAGGTCTCCATTTTACTATTCGT	381
FT	ACT_SITE 350 350	Db	76	GlyLysGlyValHisMetAsnProTyrSerTrpAsnGlnValAlaAsnIleLeuPheLeu	95
FT	ACT_SITE 403 403	QY	382	TCCAGCGATTGGGAGTGGCGTCTTTCATATAGTATGATACGGTTGATGGTCCATTAAACCT	441
FT	DISULFID 56 313	Db	96	AspSerProValGlyValGlyTyrSerTyrSerAsnThr-----	108
FT	DISULFID 217 229	QY	442	GTAAGTGGGGTGGTGGAAAATTCGAGCTTTGCAGGAGTTTCAGGCGCGGTACCAACCACT	501
FT	DISULFID 253 281	Db	109	SerAlaAspIleLeuSerAsn-----GlyAspGluArgThrAla	121
FT	CARBOHYD 107 107	QY	502	GATGCCACTCTCATGATACCTACCAATCTCCGCGAGAGCGCGTGGGAGATCGTCAA	561
SQ	SEQUENCE 436 AA; 48952 MW; E0F82D97E0C34DC9 CRC64;	Db	122	LysAspSerLeuValPheLeuThrLys-----Trp-----LeuGlu	133
		QY	562	GGATTCCTTAGTGGACTACCTAGCTTGGACTTAGGGTGCAGTCTAAGAGCTACGTCTA	621
		Db	134	ArgPhe-----ProGlnTyrLysGluArgGluPheTyrLeu	145
		QY	622	TGGACGGAGAGCTATGGAGGGCAGCTATGGTCTCGCATCTTCAATCATTTTTACGAGAG	681
		Db	146	ThrGlyGluSerTyrAlaGlyHisTyrValProGlnLeuAlaGlnAlaIleLysArgHis	165
		QY	682	AATGAGAGAATTCGCAACGGTAGTGTTAATGGTGTGTACGTTAATTTCACTCTCTGGGA	741
		Db	166	HisGluAlaThrGlyAspLysSerIle-----AsnLeuLysGlyTyrMet	180
		QY	742	ATTATTAAGCGCATCATCGACGAGGCGATCCAGGCCCTTACTACCTCAATTCGCTGG	801
		Db	161	ValGlyAsnAlaLeuThrAspAspPhe-----HisAspHisTyrGlyIlePheGlnTyr	198
		QY	802	AACAATACCTACGGTATCAGGCTGTCAACGAGACCGCTTACAACTACAAAGATTTCGC	861
		Db	199	MetTrpThrThrGlyLeu-----IleSerAspGlnThrTyrLysLeuLeuAsnIlePhe	216

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: U28730; AAA68259.1; -  
 PIR: T16606; T16606.  
 HSP: P10619; LIVY.  
 MEROPS: S10.002; -.  
 WormPep: K10B2.2; CE02009.  
 InterPro: IPR000379; Ser\_estr.site.  
 DR InterPro: IPR001563; Serine.carbpept.  
 DR Pfam: PF00450; serine.carbpept; 1.  
 DR PRINTS: PR00724; CRBOXYPTASEC.  
 DR ProDom: PD001189; Serine.carbpept; 2.  
 DR PROSITE: PS00131; CARBOXYPEPT\_SER\_SER; 1.  
 DR PROSITE: PS00560; CARBOXYPEPT\_SER\_HIS; 1.  
 KW Hypothetical protein; Hydrolase; Carboxypeptidase; Glycoprotein;  
 KW signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 470 POTATIVE SERINE CARBOXYPEPTIDASE K10B2.2.  
 FT ACT\_SITE 169 169 BY SIMILARITY.  
 FT ACT\_SITE 380 380 BY SIMILARITY.  
 FT ACT\_SITE 441 441 BY SIMILARITY.  
 FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 470 AA: 53158 MW: CCG2DACB75EF30FC CRC64;

## Alignment Scores:

Pred. No.: 5,73e-17 Length: 470  
 Score: 352.50 Matches: 131  
 Percent Similarity: 38.43% Conservative: 60  
 Best Local Similarity: 26.36% Mismatches: 179  
 Query Match: 12.14% Indels: 127  
 DB: 1 Gaps: 20

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x YSS2\_CABEL (1-470)

QY 136 ACCCGGGTGCAATCCCTACTCTGGATATGTCGACACCTCTCCCGAGTCCCATACCTTC 195  
 DB 35 ThrProAspPhePheHisTyrSerGlyTyrLeuArgAlaThrThrAspTyrLeuHis 54  
 QY 196 TTCCTGGTCTTCGAGCCAGACATACCCGAACTGCACCTATCATTGTGTTGAAT 255  
 DB 55 TyrTrpLeuThrGluSerSerArgAlaProThrGlnAspProLeuValLeuTriLeuAsn 74  
 QY 256 GGTGGCCCTGGAAGCATCTCTTCATCGCTCTTCGAGAGTGGGCCCTGCCATGTC 315  
 DB 75 GlyGlyProGlyCysSerSerLeuAspGlyLeuLeuGluGluGlyProPheHisVal 94  
 QY 316 AATTCGACTTTGTATGAC---TACATCAACCTCCTGCTGGAGACGAGGTCTCCAATTIA 372  
 DB 95 LysAspPheGlyAsnSerIleTyrTyrAsnGluTyrAlaThrPsnLysPheAlaAsnVal 114  
 QY 373 CTATTCCTGTCGCCAGCATGGGAGTCGGCTTTTCATATAGTACAGTGTGGGTC 432  
 DB 115 LeuPheLeuGluSerProAlaGlyValGlyTyrSerTyrSerThrAsnPheAsnLeuThr 134  
 QY 433 ATTAAACCTGTACTGGGGTGGCTCGAAAATTCGAGCTTTCGAGAGITTCAGGCCGGTAC 492  
 DB 135 ValSerAspAspGluValSerLeuHisAsn----- 144  
 QY 493 CCAACCATTTGATCCACTCTGATGATACCTACCAATCTTCCGAGAGCGGCTGGGAG 552  
 DB 145 -----TyrMet 146  
 QY 553 ATCTGTCAAGGATTCCTTAGTGGACTACCTAGCTTGACTCTCAGGTGCATCAAGGAC 612  
 DB 147 AlaLeuLeuAspPheLeuSerLysPheProGluTyrLysGlyArg-----Asp 162  
 QY 613 TTCAGTCTATGGACGGAGCTATGGAGGACCTATGGCTGCTGCATCTCTCAATCAITTT 672  
 DB 163 PheTrpIleThrGlyGluSerTyrAlaGlyValTyrIleProThrLeu----- 178

QY 673 TACGACGAGAAATGAGAGAAATGCCAACGGTAGTCTTAATGTTGTGTACGTTAATTCAAC 732  
 DB 179 -----AlaValArgIleLeuAsnAspLysLysAsnPhePro-----AsnPheLys 193  
 QY 733 TCCTCTGGAAATTTAATACGGCATCATC-----GACGAGGAGATCCAGGCCCT 780  
 DB 194 GlyValAlaIleGlyAsnGlyAlaLeuAsnPheProAsnAsnTyrAsnThrMetValPro 213  
 QY 781 TACTACCTCGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGACCGCIC 840  
 DB 214 PheTyr-----TyrTyrHisAlaLeuValArgAspLeu 225  
 QY 841 TACAACACTACATGAAGTTTGCCAAACAAATGTTGCCAGGATTTGATTTCCACC 900  
 DB 226 TyrAsn-----AspIleAlaArgAsnCysCysAsnAsnAsnIleGlyThr 240  
 QY 901 TGCACACAGACAACCCGACCGCATFAGCTAGCTACGCCCTCTGCGCCGAGCCACCAAC 960  
 DB 241 CysAsp-----IleTyrSerLysPhePheAspProAsn--- 251  
 QY 961 ATGTGACGAGCAATGTGTGAGGGCCATACCTACGCCCTTCTGCTGCTGTGTGTATGAT 1020  
 DB 252 ---CysArgAspLysVal-----IleAsnAlaLeuAspGlyThrAsnGluLeuAsn 267  
 QY 1021 AITCGGCATCCATATGATGATGACCGCAGCTCCGCCAAGTTATTACAAAC 1065  
 DB 268 MetTyrAsnLeuTyrAsp-----ValCysTyrTyrAsnProThrThrAsnLeu 283  
 QY 1066 ---AAATTTTCGGCAAGGACCTGTCATGCGGCTATCGGGCTC----- 1107  
 DB 284 LysLysAlaPheIleGluArgGlnMetArgIleAlaValGlyLeuProAlaArgLysHis 303  
 QY 1108 AACATCAACTACACAC-----CAGTCCAATAATGACGCTCTACTAC----- 1146  
 DB 304 AsnAlaAlaThrValProLeuCysAlaGlnThrAsnAsnThrHisValTyrLeuAsn 323  
 QY 1147 -----GCTTCCAGCAACA 1161  
 DB 324 ArgAlaAspValArgLysSerLeuHisIleProSerSerLeuProAlaThrPgluGlyCys 343  
 QY 1162 GSCGACTTGTCTGCGCCCAACITC-----ATCGAAGACCTCGAG 1200  
 DB 344 SerAspGlnValGlyLysAsnTyrValThrHisPheAsnValIleProGluPheGln 363  
 QY 1201 GAGATCTCTGCTCTCCCGTGTCTCCCTCATCTATGCGGACGCGATTCATCTGCG 1260  
 DB 364 ThrMetIleAlaAlaGlyIleLysIleLeuValTyrAsnGlyAspValAspThrAlaCys 383  
 QY 1261 AACTGGTTCGGGTCGAGCGGCTTTCCTGCTGCGAACTACICCCAGCCGCCAG--- 1317  
 DB 384 AsnSerIleMetAsnGlnGlnPheLeuThrSerLeuAsnLeuThrValLeuGlyGluGln 403  
 QY 1318 -----TTCGAGCGCAGGATACACGCCCTGCAAGCTCAACGCGCTC 1359  
 DB 404 GluLysValAsnGluAlaIrrHisTyrSerGlyGlnThrGlyThrAlaValAlaGlyPhe 423  
 QY 1360 GAGTAIGGGGAACTCGCGAGTATGTAATTTCTCTCACTCGCTGCTATGAGGCGAGC 1419  
 DB 424 -----GlnThrLysPheAlaGlyAsnValAspPheLeuThrValArgLysGly 440  
 QY 1420 CAIGAAGTCCCATCTACACGCCCATGCCCTCCCTGCAA---TTGTTTAAAC 1467  
 DB 441 HisPheValProGluAspLysProLysGluSerGlnGlnMetIlePheAsn 457

## RESULT 15

NE31\_NAEFO  
 ID NE31\_NAEFO STANDARD; PRT; 482 AA.  
 AC P42661;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Virulence-related protein NF314 (EC 3.4.16.-).

OS Naegleria fowleri.  
OC Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.  
OX NCBI\_TaxID=5763;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=LEE;  
RC MEDLINE=92267659; PubMed=1587609;  
RX Hu W.-N., Kopachik W., Band R.N.;  
RA "Cloning and characterization of transcripts showing  
RT virulence-related gene expression in Naegleria fowleri.";  
RL Infect. Immun. 60:2418-2424(1992).  
CC -I- FUNCTION: MAY BE REQUIRED BUT IS NOT SUFFICIENT FOR INCREASED  
CC VIRULENCE.  
CC -I- INDUCTION: BY GROWTH ON MAMMALIAN CELLS.  
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

-----

EMBL; M883397; AAA29384.1; .  
DR PIR: A43828; A43828.  
DR HSP: P10619; LIVY.  
DR MEROPS; S10.UPW.  
DR InterPro; IPR000379; Ser\_estrs\_site.  
DR InterPro; IPR001563; Serine\_carbpept.  
DR Pfam; PF00450; serine\_carbpept; 1.  
DR PRINTS; PR00724; CRBOXYPTASEC.  
DR PRODOM; PD001189; Serine\_carbpept; 2.  
DR PROSITE; PS00131; CARBOXYPEPT\_SER\_SER; 1.  
DR PROSITE; PS00560; CARBOXYPEPT\_SER\_HIS; 1.  
DR Hydrolase; Carboxypeptidase.  
KW ACT\_SITE 163  
FT ACT\_SITE 399 399 BY SIMILARITY.  
FT ACT\_SITE 459 459 BY SIMILARITY.  
SQ SEQUENCE 482 AA; 53848 MW; 0B83049C1A8A0908 CRC64;

Alignment Scores:  
Pred. No.: 7,34e-17 Length: 482  
Score: 351.00 Matches: 139  
Percent Similarity: 35.19% Conservative: 57  
Best Local Similarity: 24.96% Mismatches: 165  
Query Match: 12.09% Indels: 196  
DB: 1 Gaps: 23

US-09-712-338-I\_COPY\_55\_1662 (1-1608) x NF31\_NAFPO (1-482)

QY 109 CCGGGGAGAGGGGTCTGCAGACTACCCGGGTGCAATCTACTCTGGATATGTC 168  
||||| ||| ||||| |  
DB 20 ProglyLeuSerGly-----AsnIleGlyVallySerTyrThrGlyTyrLeu 35  
  
QY 169 -----GACACTCTCCGAGTCCCATACTTCTTCGTCTTCGAAGCCAGACATAAC 222  
||| :|: :|: ||||| |  
DB 36 LeuAlaAsnAlaThrArgGlyArgTyrLeuPheTyrTrpPheLeuUserMetArgAsn 55  
  
QY 223 CCAGAATCGCACCTCATCATTTGGTGATGGTGGCCCTGGAGGCGATTCTTTGATC 282  
||| ||||: :|: ||||| |  
DB 56 ProSerGlnAspProLeuValMetIrrpThrAsnGlyGlyProGlyCysSerSerLeuGly 75  
  
QY 283 GGTCTCTTCGAAGATGGGCCCTGCCATGCAATTCGACACTTTTGATGACATC --- 339  
||| ||| ||| ||||| |  
DB 76 GlyGluAlaSerGluHisGlyLeuPheLeuValAsnAlaAspGlyAlaThrIleThrArg 95  
  
QY 340 AACCTCACTCTCGTGAAGACGAGTCTCCAATTTACTATTCTCTCCAGCCATTGGGAGTC 399  
||||| :|: ||||| |  
DB 96 AsnProTyrSerIrrpAsnArgValSerAsnIleLeuTyrIleGluGlnProValGlyVal 115  
  
QY 400 GGCTTTTCATATAGTAGTACGGTTGATGGGTCCATTAACTCGTGTACTGGGTGCTGCGAA 459  
||||| :|: ||||| |

```
Oy 1264 TGGTTCGGCGTCAAGCGGTTTCCTCGCTCGGAACACTCCCAAGCGCGCCAGTCCGA 1323
    ||| |||||
Db 404 GlyLeuGlyThrGlnAlaAlaIleAspLysLeuGlnLeuGlnGluThrSerSerIrpArg 423
    : : : : :
Oy 1324 AGCGCAGGTACACGCCCTGAAAGTCAACGGCGTCGAGTATGGGAAACTCGCGAGTAT 1383
    : : : : :
Db 424 ThrTrpGluPheAspSer---AlaLeuGlyThrValValGlyGlyTyrIleArgLysPhe 442
    : : : : :
Oy 1384 -----GGTAATTCTCTCACTCGCGTCTATGAGCGCAGGCCCATGAAGTCCCA 1431
    : : : : :
Db 443 GluLysSerGlyLysGlyLeuThrPheLeuThrValArgGlyAlaGlyHisMetValPro 462
    : : : : :
Oy 1432 TACTACAGCCCATCGCTCCCTGCAATTGTTTAACCGGACTATCTCGGT 1482
    : : : : :
Db 463 LeuValLysProAspSerAlaPheTyrMetPheLysAsnPheIleAspGly 479
```

Search completed: September 16, 2003, 18:43:07  
Job time : 58 secs

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 16, 2003, 18:14:18 ; Search time 111.5 Seconds  
(without alignments)  
7443.024 Million cell updates/sec

Title: US-09-712-338-1\_copy\_55\_1662

Perfect score: 2904  
Sequence: 1 cttccagggaagtacacggc.....gcattgtccagtgtgtgatg 1508

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=framet\_n2p.model -DBV=xlp  
-Q=/cgn2.1/USPTO.spool.p/US09712338/runat\_16092003\_144325\_14356/app\_query.fasta\_1.1799  
-DB=SPREMBL\_23 -QFWT=fastan -SURFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09712338.ecgn\_1.123.0runat\_16092003\_144325\_14356 -NCPD=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONSGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPREMBL\_23:.\*  
1: sp\_archaea:.\*  
2: sp\_bacteria:.\*  
3: sp\_fungi:.\*  
4: sp\_human:.\*  
5: sp\_invertebrate:.\*  
6: sp\_mammal:.\*  
7: sp\_mhc:.\*  
8: sp\_organelle:.\*  
9: sp\_phage:.\*  
10: sp\_plant:.\*  
11: sp\_rodent:.\*  
12: sp\_virus:.\*  
13: sp\_vertebrate:.\*  
14: sp\_unclassified:.\*  
15: sp\_virus:.\*  
16: sp\_bacteriap:.\*  
17: sp\_archaeap:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2875	99.0	555	3 Q96VZ9	Q96vz9 aspergillus

2	746.5	25.7	460	3 Q12569	Q12569 absidia zyc
3	601	20.7	552	3 Q96VC4	Q96vc4 emericella
4	596.5	20.5	541	3 Q94152	Q94152 pichia angu
5	595.5	20.5	537	3 Q14414	Q14414 pichia angu
6	479	16.5	507	10 Q8VWQ0	Q8vwq0 gossypium h
7	473	16.3	429	10 Q8GVU1	Q8gvu1 oryza sativ
8	466.5	16.1	505	10 Q9FFB0	Q9ffb0 arabidopsis
9	461	15.9	501	10 Q9XH61	Q9xh61 matricaria
10	460	15.8	508	10 Q8L6A7	Q8l6a7 theobroma c
11	439	15.1	510	10 Q9LXH4	Q9lxl4 arabidopsis
12	438	15.1	510	10 Q932C3	Q932c3 arabidopsis
13	413.5	14.2	2105	5 Q17679	Q17679 caenorhabdi
14	406.5	14.0	499	10 Q9FMX9	Q9fmx9 arabidopsis
15	385	13.3	490	10 Q9FYF7	Q9fyf7 oryza sativ
16	383	13.2	482	10 Q9SV04	Q9sv04 arabidopsis
17	376.5	13.0	524	10 Q8GTK2	Q8gtk2 oryza sativ
18	363	12.5	480	10 Q9LEV1	Q9lev1 arabidopsis
19	362.5	12.5	487	10 Q9SV03	Q9sv03 arabidopsis
20	360	12.4	437	10 Q9FWG1	Q9fwg1 oryza sativ
21	360	12.4	437	10 Q9FWG1	Q9fwg1 oryza sativ
22	358	12.4	482	10 Q9FRJ0	Q9frj0 oryza sativ
23	357.5	12.3	479	10 Q94907	Q94907 arabidopsis
24	355.5	12.2	360	10 Q9M450	Q9m450 cicer ariet
25	349	12.0	474	11 Q9D2D1	Q9d2d1 mus musculu
26	348	12.0	452	10 Q9ZQ00	Q9zq00 arabidopsis
27	344	11.8	465	10 Q04084	Q04084 arabidopsis
28	344	11.8	494	10 Q9FEU4	Q9feu4 pisum sativ
29	344	11.8	501	10 Q9SV02	Q9sv02 arabidopsis
30	342.5	11.8	471	10 Q8LY00	Q8ly00 arabidopsis
31	342.5	11.8	473	10 Q9SCA9	Q9scs9 arabidopsis
32	342.5	11.8	669	10 Q8LPY6	Q8lpy6 oryza sativ
33	340	11.7	512	5 Q76725	Q76725 caenorhabdi
34	337.5	11.6	425	10 Q65568	Q65568 arabidopsis
35	333.5	11.5	510	3 Q60123	Q60123 schizosacch
36	333	11.5	487	10 Q9MUG3	Q9mug3 arabidopsis
37	332	11.4	465	10 Q9M099	Q9m099 arabidopsis
38	331	11.4	497	10 Q8L7B2	Q8l7b2 arabidopsis
39	330.5	11.4	472	10 Q9LSM9	Q9lsm9 arabidopsis
40	330.5	11.4	474	10 Q82228	Q82228 arabidopsis
41	330	11.4	445	5 Q814E3	Q814e3 caenorhabdi
42	330	11.4	1203	5 Q45916	Q45916 caenorhabdi
43	329.5	11.3	548	4 Q9BR08	Q9br08 homo sapien
44	328.5	11.3	434	11 Q9D3S9	Q9d3s9 mus musculu
45	328	11.3	504	10 Q9LSV8	Q9lsv8 arabidopsis

ALIGNMENTS

RESULT 1  
Q96VZ9 ID Q96VZ9 PRELIMINARY; PRT; 555 AA.  
AC Q96VZ9;  
DI 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Carboxypeptidase S1 precursor (EC 3.4.16.6).  
GN CPl.  
OS Aspergillus oryzae.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5062;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TK3;  
RA van den Broek P.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF394242; AKK7166.1; -  
DR InterPro; IPR001563; Serine\_carbpept.  
DR Pfam; PF00450; serine\_carbpept; 1.  
DR PRINTS; PR00724; CRBOXYPTASEC.  
DR ProDom; PD001189; Serine\_carbpept; 1.  
DR PROSITE; PS00131; CARBOXYPEPT\_SER\_SER; 1.  
KW Carboxypeptidase; Hydrolase; Signal.

FT SIGNAL 1 18 POTENTIAL.  
SQ SEQUENCE 555 AA; 61168 MW; 456B63B0CEB55222 CRC64;

Alignment Scores:  
Pred. No.: 5,85e-198 Length: 555  
Score: 2875.00 Matches: 535  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.81% Mismatches: 0  
Query Match: 99.00% Indels: 0  
DB: 3 Gaps: 0

US-09-712-338-1\_copy\_55\_1662 (1-1608) x Q96VZ9 (1-555)

QY 1 CTTCCAGGAGTACACCGGCGTCCGCTGCTAGAACAGCTACCCAAAGAACCCACCGGG 60  
DB 19 LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly 38  
QY 61 GTCAGAGCTTACAAACCGCAACATGTCACATCCGGGTACAAAGAACCCGGGGCAGAG 120  
DB 39 ValLysThrLeuThrThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGlu 58  
QY 121 GCGCTGCGAGACTACCGGGGTGCAAAATCTACTCTGGATATGTCACACCTCTCCC 180  
DB 59 GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 78  
QY 181 GAGTCCCATACCTTCTCTGGTCTTCGAAAGCCAGACATAACCCAGAAACTGCACCTATC 240  
DB 79 GluSerHisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAlaProIle 98  
QY 241 ACATTTGGTGTGAATGTCGCCCTCGAAGCGATTCTTTGATCGGTCTCTTCGAAAGATTG 300  
DB 99 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeu 118  
QY 301 GGCCTTGCCATGTCATTCGACTTTTGATGACTATCATCAACCTCAGCTGCGAAGCAG 360  
DB 119 GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGlu 138  
QY 361 GTCCTCAATTTACTATCTCTCCAGCCATCGGAGTCGGCTTTTATATAGTATGATACG 420  
DB 139 ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrAsnAspThr 158  
QY 421 GTTATGGGGTCCATTAACCTGTAACGCGGTGCTCGAAATTCGAGCTTTGAGGAGTT 480  
DB 159 ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal 178  
QY 481 CAGGCGCGTACCAACCATGATGCCACTCTGATCGATACCAATCTTCCGCGCAGAG 540  
DB 179 GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGlu 198  
QY 541 GCGCTTGGGAGATCTCGAAGGATCTTCTTGTGAGTACTAGCTTGGACTCTAGGGTG 600  
DB 199 AlaAlaThrGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal 218  
QY 601 CAGCTAAGGACTTCAGTCTATGACGAGAGCTATGAGGCGCATATGTCCTGCTATC 660  
DB 219 GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 238  
QY 661 TTCATCATTTTACGACCAAGTACAGAGATTCGCAACGCTAGTGTATGTTGTTTCAG 720  
DB 239 PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 258  
QY 721 CTTAATTTCAACTCTCTGGGAATTTAATACGGCATCATPCGAGCGGATCCAGGCCCT 780  
DB 259 LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaPro 278  
QY 781 TACATACCTGAATTCGCTGTGACAAATACCTACGCTATCAGGCTCTCAACGAGACCTC 840  
DB 279 TyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrVal 298  
QY 841 TACAACTACATCAAGTTTGCAACCAATGCCAATGTTGCGAGATTTCATTTCCACC 900  
DB 299 TyrAsnTyrMeLysPheAlaAsnGlnMetProAsnGlyCysGluAspLeuIleSerThr 318

QY 901 TCGAAACAGACAAACCGCACCCGATAGCTAGTACGCCCTCTGCGCGAAGCCACCAAC 960  
DB 319 CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn 338  
QY 961 ATGTGCGAGGACAAATGTTGAGGGGCCATACCTAGCCCTTTGCTGGTCTGGTGTATGAT 1020  
DB 339 MetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAsp 358  
QY 1021 ATTCGGCATCCATATGATGACCGGACTCCGCCAAGTATTATACAAATAATTCGGCAAG 1080  
DB 359 IleArgHisProTyrAspAspProThrProSerTyrTyrAsnLysPheLeuAlaLys 378  
QY 1081 GACTCTGTCATGACGCTATTCGCGCTCAACATCACTACACCCAGCTCCATATATGAGTC 1140  
DB 379 AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal 398  
QY 1141 TACTAGCTTCCAGCAAAACAGCGCTTGTCTGCGCCCACTTCCATCATTCGAAAGACCTCGAG 1200  
DB 399 TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu 418  
QY 1201 GAGATCTTGTCTCTCCCGTGTCTCTCCCTCATCTATGCGGACGCGGATTCATCTGC 1260  
DB 419 GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys 438  
QY 1261 AACTGGTTCGCGGTGAGCGCTTCCCTCGCTGCGAAGTACTCCCAAGCGCCGACGTTTC 1320  
DB 439 AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGlnPhe 458  
QY 1321 CGAAGCGCAGGGTACACGCGCCCTGAAAGTCAAGCGCTCGAGTATGGGAAACTCGCGAG 1380  
DB 459 ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu 478  
QY 1381 TATGGTAATTTCTTCCTCCTCCTGCTATGAGCGCGCATGAGTCCCATACCTACACAG 1440  
DB 479 TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln 498  
QY 1441 CCATCCCTCCCTGCAATTTTACCGGACTATCTTCGTTGGGATATCGCAGAGGCG 1500  
DB 499 ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGly 518  
QY 1501 CAGAAGAGATCTCGCCGACGTACAGACGAAATGAAAGCGGTACAGCTACGCTACGATACAG 1560  
DB 519 GlnLysLysIleTyrProSerTyrIleThrAsnGlyThrAlaThrAlaThrHisThrGln 538  
QY 1561 TCGTCCGTCGCGCTCGCTACGCTACGAGTACGAGTATGCGAGTGTGTATG 1608  
DB 539 SerSerValProLeuProThrAlaThrSerMetSerSerValGlyMet 554

RESULT 2

Q12569 PRELIMINARY; PRT; 460 AA.  
ID Q12569;  
AC Q12569;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Prepro-carboxypeptidase 2 (EC 3.4.16.5).  
GN spcz.  
OS Absidia zychoae.  
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;  
OC Absidia.  
OX NCBI\_TaxID=36079;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRIC 1199;  
RX MEDLINE=95308538; PubMed=7788719;  
RA Lee B.R., Takeuchi M., Kobayashi Y.;  
RI "Molecular cloning and sequence analysis of the scp2 gene encoding the  
RL serine carboxypeptidase of Absidia zychoae."  
RL Curr. Genet. 27:159-165(1995).  
DR EMBL; D16519; BAA03966.1; -.  
DR HSSP; P00729; IYSC.  
DR MEROPS; S10.010; -.  
DR InterPro; IPR001563; Serine\_carbpept.

InterPro: IPR000379; Ser\_estrs\_site.  
Pfam: PF00450; serine\_carbpept; 1.  
DR PRINTS; PR00724; CRBOXIPTASEC.  
DR ProDom; PD001189; Serine\_carbpept; 1.  
KW Carboxypeptidase; Hydrolase.  
FT CHAIN 52 460 MATURE CARBOXYPEPTIDASE Z.  
SQ SEQUENCE 460 AA; 50825 MW; CE3A410764548BB6 CRC64;

Alignment Scores:  
Pred. No.: 3-65e-45 Length: 460  
Score: 746.50 Matches: 175  
Percent Similarity: 50.31% Conservative: 71  
Best Local Similarity: 35.79% Mismatches: 182  
Query Match: 25.71% Indels: 61  
DB: 3 Gaps: 14

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x Q12569 (1-460)

QY	16	CGCGCGTCGTCGGTAGAGACAGCTACCCAGAAACCCACCGGGGTCAAGACTCTTACA	7	...
Db	29	ProMetSerHisValLeuArgArgGlnAspGlyAsnAspThr	...	---Ser 43
QY	76	ACCGCAAAACAATGTACCATCCGTTACAGAAACCCGGGCGAGGCGGTCTCGGAGACT	135	...
Db	44	SerGlyAsnThrThrGlnLeuIysThrSerPro	...	---LysLeuCysAsp--- 59
QY	136	ACCCCGGGTGCAAACTCTACTCTGGATATCTGCACACCTCTCCGAGTCCCATACCTIC	195	...
Db	60	---ProAspValLysGlnTyrSerGlyTyrLeuAspAlaAlaAsnAspGluHisTyrPhe	78	...
QY	196	TTCTGGTCTTCGAAGCCAGACATACCCAGAAACTGCATCATCATTTGTGGTTCGAAT	255	...
Db	79	PheTrpPheGluSerLysAsnAspProLysAsnAspProLeuThrIleTyrLeuAsn	98	...
QY	256	GTGGCCCTGGAAGCGAATTTTCATGGTCTCTTCCAGAGTTGGCCCTTCGCATGTC	315	...
Db	99	GlyGlyProGlyCysSerSerLeuIleGlyTrpGluGluLeuGlyProCysGlnGln	118	...
QY	316	AATCGACTTTTGATGATACATCAACCCCTCACTCGTGGAAACGAGGTCTCCAAATTTACTA	375	...
Db	119	AsnGlySer	...	---AlaAsnProHisSerTrpHisSerSerAsnMetLeu 134
QY	376	TTCCGTGCCAGCCATGGGAGTCGGCTTTTTCATATAGTCATACGTTGATGGTCCATT	435	...
Db	135	PhePheAspGlnProAspGlyValGlyPheSerTyr	...	--- 146
QY	436	AACCCTGTAACTGGGGTCGTGAAATTCGAGCTTTCGAGGAGTTCAGGCGCGGTACCCA	495	...
Db	147	---	---	---GlyLysGln--- 149
QY	496	ACCAITGATGCCACTCTGATGATACCTACCAANTCTTCCGCAGAGCGCGTGGGAGATC	555	...
Db	150	---	---	---ThrValSerThrThrGluAspAlaAlaGluArgAlaTrpThrPhe 164
QY	556	CTGCAAGGATTCCTTAGTGACTACTAGTTGGACTCTAGGTGCAGTCTAAGGACTTC	615	...
Db	165	LeuGlnAlaPheTyrGluThrPhePro	...	---GlnTyrSerLysLeuAspVal 180
QY	616	AGTCTATGCGGAGAGCTATGAGGGCACTATGGTCTCGTCACTTTCATCATTTTAC	675	...
Db	181	HisTyrPheGlyGluSerTyrGlyGlyHisTyrIleProGlyPheAlaSerHisValVal	200	...
QY	676	GAGCAGAAATGAGAAATGGCAACGGTAGTGTAAATGGTGTTCAGCTTAATTCACACTC	735	...
Db	201	AspMetAsnLysLysValGlnSerGlyGluGluLysValValProLeuLysSer	220	...
QY	736	CTGGCAATTAATACGGCATCATCAGGAGGCGATCCAGGCCCTTACTACCTGAATTC	795	...
Db	221	IleGlyValGlyAsnGlyPheIleAspAlaValIleGlnTyrLysSerTyrProLysMet	240	...
QY	796	GCTGTGAACAATACTACGGTATCAGGCTGTCAACGAGACCGTCTACAACTACATGAAG	855	...
Db	241	ThrCysHisSerThrTyrProAlaValLeuSerGluGlu	...	---GlyCysAspLys 257

DR PROSITE: PS00131; CARBOXYPEPT\_SER\_SER; 1.  
SQ SEQUENCE 552 AA; 62065 MW; F67CF39FDBF7D761 CRC64;

## Alignment Scores:

Pred. No.: 1.09e-34 Length: 552  
Score: 601.00 Matches: 154  
Percent Similarity: 47.34% Conservativeness: 77  
Best Local Similarity: 31.56% Mismatches: 175  
Query Match: 20.70% Indels: 82  
DB: 3 Gaps: 19

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x Q96VC4 (1-552)

```
QY 79 GCAACAATGTCACCATCGGTACAGGAACCCGGGCGAGGGCGTTCGCGAGACTACC 138
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
127 AlatyAspLeuArgLleuLysLysThrAspProSerSerLeuGlyIle-----Asp 143
QY 139 CCGGTGTCAATCTACTCTCGATATGTCGACACTCTCC-----GAGTCCCATCTTC 195
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
144 ProAspValLysGlnThrGlyLysLeuAspAspAsnGluAsnAspLysHisLeuPhe 163
QY 196 TTCTGTTCTTCGACCCAGACATACCCAGAACTGCACCTATCATCTGTTGTTGAAT 255
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 TyrTrpPheGluSerArgAsnAspProLysAsnAspProValValLeuTrpLeuAsn 183
QY 256 GGTGGCCCTGGAGCGATCTTTGATCGGTCTCTTCGAAGAGCTGGGCCCTTCCCATGTC 315
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetClnLeuGlyProSerSerIle 203
QY 316 AATTCGACTTTTGATGACATCAACCTCACTCTCTGGAACGAGCTCTCCATTTACTA 375
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
204 AspGluAsnLleLysProValTyrAsnProTyrAlaTrpAsnSerAsnAlaSerValIle 223
QY 376 TTCTGTTCCCGACCATTTGGAGCTCGCTTTTCATATAGTGATACGGTGTGATGGTCCATT 435
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
224 PheLeuAspGlnProValAsnValGlyTyrSerTyrSer----- 236
QY 436 AACCTGTAACTGGGGCTGCGAAAATTCGAGCTTTCGAGGAGTTCAGGCGCGTGACCCA 495
Db 236 ----- 236
QY 496 ACCATGATGCCACTCTGATCGATACATACCAATCTTCCGCCAGAGCGCGTTCGGAGATC 555
Db |||:||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
237 -----GlySerThrValSerAspThrVal---AlaAlaGlyLysAspValTyrAlaLeu 253
QY 556 CTGCAAGGATTCCTTAGTGACTACCTAGCTTGGACTCTAGGTGCGAGTCTAAGGACTTC 615
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
254 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaGluGlnAspPhe 269
QY 616 AGTCTATGGAGGAGCTATGGAGGCACTAIGTCTGCTCATCTTCAATCAATTTTAC 675
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
270 HisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheThrSerGluIleLeu 289
QY 676 GAGCAGATGAGAGATTCGCCAGGTAGTGTGTAATGGTGTTCAGCTTAATTTCAACTCT 735
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
290 SerHisGlnLysArg-----AsnIleAsnLeuLysSer 300
QY 736 CTGGGAATTTATACGSCATCATCGACGAGCGCATCCAGCCCTTACTACCTCGAATTC 795
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 ValIleGlnGlyAsnGlyLeuThrAspGlyLeuThrGlnTyrGluTyrTyrArgProMet 320
QY 796 GCTGTGAACAATACCTACGCTATCAAGGCTGTCAACGAGACCGCTCTACAACATCAAG 855
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
321 AlaCysGlyGluGly---GlyTyrProAlaValLeuAspGlu-----Ser 334
QY 856 TTGCGCCAAACCAATGCCAATGGT-----TGCCAGGATTTGATTTCCACCTGAAA 906
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
335 SerCysArgSerMetAspAsnAlaLeuGlyArgCysGlnSerMetIleGluSerCysTyr 354
QY 907 CAGACAAACCGCACCGCATTAGCTAGCTACGCCCTCTGGCGGAGGACCAACATCGTC 966
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
355 AsnSerGluSerAla-----TrpValCysValProAlaSerIleTyrCys 369
```

```
QY 967 AGGACAATGTTGAGGGGCCATACTACGCGCTTTCCTGGTCTGGTCTATGATATCGG 1026
Db |||:||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
370 AsnAsnAlaLeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArg 388
QY 1027 CATCATATGATGACCCCGACT-----CCGCCAAGTTATTACAAACAATTT 1071
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
389 GlyLysCysGluAspGluSerAsnLeuCysTyrLysGlyMetGlyTyrValSerGluTyr 408
QY 1072 CTGGCAAGAGGACTCTGTCTGATGACGCTATCGGCGTCAACATCAAC---TACACCCAGTCC 1128
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
409 LeuAsnLysProGluValArgAlaAlaValGlyAlaGluValAspGlyTyrAspSerCys 428
QY 1129 AATAATGACGCTTACTACGCTTTCACAGCAAAACAGGCGACTTTCCTGGGCC-----AAC 1182
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
429 AsnAspIleAsnArgAsnPheLeuPheHisGlyAspTrpMetLysProTyrHisArg 448
QY 1183 TTCATGAAGACCTCGAGGAGATCTTGTCTCCCGTGGTCTCTCCCTCATCTAT--- 1239
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
449 LeuValProGlyIleLeuGlu-----GlnIleProVal-----LeuIleTyrAla 463
QY 1240 GCGAGCGCGATATACATCTGCAACCTGTCGGGGTTCAGGCCCTTCCCTCGCTGCGAAC 1299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
464 GlyAspAlaAspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGlu 483
QY 1300 TACTCCCAACCCCGCTCCGAGTCCGAGCGAGGTACAGCCCGCTGAAGTC----- 1350
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
484 TrpProGlyHisLysGluPheAlaAlaProMetGluAspLeuLysIleValAspAsn 503
QY 1351 -----AACGGCTCGAGTATGGGAAACTCGGAGTATGATGTAATTTCTCTCACPTCGC 1404
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
504 GluHisThrGlyLysLysIleGlyGlnIleLysThrHisGlyAsnPheThrPheMetArg 523
QY 1405 GTCTATGAGCAGCGCCATCAAGTCCCATCTACAGCCCATCGCCTCCCTGCAATTTT 1464
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
524 LeuTyrGlyGlyGlyHisMetValProMetAspGlnProGluAlaSerLeuGluPhePhe 543
QY 1465 AACGGAGCTATCTTCGGT---TGG 1485
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
544 AsnArgTrpLeuGlyGlyGluTrp 551
RESULT 4
O94152 PRELIMINARY; PRT; 541 AA.
AC O94152;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase Y.
GN CPT.
OS Pichia angusta (Yeast) (Hansenula polymorpha).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4905;
EN |||
RP SEQUENCE FROM N.A.
RC STRAIN=CBS4732;
RX MEDLINE=99174803; PubMed=10077185;
RA Bellu A.R., van der Klei I.J., Reehinger K.B., Yavuz M., Veenhuis M.,
RA Kiel J.A.K.W.;
ET "Characterization of the Hansenula polymorpha CPT gene encoding
ET carboxypeptidase Y";
RL Yeast 15:181-189(1999).
DR EMBL; AF085063; AAC96121.1; -.
DR HSSP; P00729; 1YSC.
DR MEROPS; S10.001; -.
DR InterPro; IPR001563; Serine_carbept.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00450; serine_carbept; 1.
DR PRINTS; PR00724; CRBOXPTASEC.
DR ProDom; PD001189; Serine_carbept; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
SQ SEQUENCE 541 AA; 60793 MW; A00CCB58526B1753 CRC64;
```

## Alignment Scores:

Pred. No.: 2,27e-34 Length: 541  
 Score: 596.50 Matches: 164  
 Percent Similarity: 46.14% Conservative: 63  
 Best Local Similarity: 33.33% Mismatches: 184  
 Query Match: 20.54% Indels: 81  
 DB: 3 Gaps: 20

US-09-712-338-1\_copy\_55\_1662 (1-1608) x 094152 (1-541)

```

QY 82 AACATGTCACATCCGCTACAAG-----CAACCGGGGAGAGGGCTGCGAGACT 135
DB 115 SerAsnTyrLysLeuArgValLysLysThrProGlyAlaLeuLeuLeuAspHisThr 134
QY 136 ACCCGGGGTGCAATCTACTCTGGATATCGACACCTCTCCGAG---TCCATACC 192
DB 135 -----ArgGlnTyrSerGlyTyrLeuAspValGluAspGluAspLysHisPhe 150
QY 193 TTCCTCTGGTCTTCGAAGCCAGACATAACCCAGAACTGCACCTATCACATTTGGTGG 252
DB 151 PheTyrIrpMetPheGluSerArgAsnAspProAlaAsnAspProValIleLeuTyrLeu 170
QY 253 AATGGTCGCTCGAAGCGATCTTTCATCGGCTCTCTCGAAGAGTGGGCCCTTGGCAT 312
DB 171 AsnGlyGlyProGlyCysSerSerLeuThrGlyMetLeuPheGluLeuGlySerAlaSer 190
QY 313 GTCAATTTCGACTTTTGTGATGACTACATCAACCTCCTCGTGGAGAGGCTCCAAATTA 372
DB 191 IleGlyProAspLeuLysProIleAsnAsnProTyrSerIrpAsnSerAsnAlaThrVal 210
QY 373 CTATTCCTGTCGCCAGCATTTGGAGTCGGCTTTTCATATAGTATGATGAGTGGTGGTCC 432
DB 211 IlePheLeuAspGlnProValAsnValGlyPheSerTyrSer----- 224
QY 433 ATTAACCCCTGAATCGGGTCGTCGAAATTCGAGCTTCGAGAGTTCAGGGCCGGTAC 492
DB 224 ----- 224
QY 493 CCAACCATTTGATGCCACTCTGATCGATACCTACCAATCTTCGCGCAGAGCGGCTGGAG 552
DB 225 -----SerLysSerValSerAsnThrValAlaAlaGlyLysAspValTyrAla 240
QY 553 ATCCGTCAGAGTTCCTTACTGGACTACCTAGCTTGGACTAGAGTGCAGTCTAAGCA 612
DB 241 PheLeuGluLeuPheTyrGlnGlnPheProHisLeu-----LeuLysAsnAsp 256
QY 613 TTCAGTCTATGAGCAGAGACTATGAGGCGCACTATGCTCCTGCATCTTCATCATTTT 672
DB 257 PheHisIleAlaGlyGluSerTyrGlyGlyHisTyrIleProValPheAlaSerGluIle 276
QY 673 TAGCAGCAGAAATGAGAAATGCCAAGGTAGTGTAAATGGTGTTCAGCTTAATTTCAAC 732
DB 277 LeuThrHisAlaAspArg-----SerPheAsnLeuThr 287
QY 733 TCCTCTGGATATTAACGGCATCATCGAGAGCGCATCCAGCGCCCTACTACCTCGAA 792
DB 288 SerValLeuIleGlyAsnGlyLeuThrAspProLeuAsnGlnTyrProPheTyrGluArg 307
QY 793 TTCGCTGTGAACAATACCTACGGTATCAAG---CGCTCAACGAGACCGTCTCACTAC 849
DB 308 MetAlaCysSerThrAspGlyTyrGluProThrLeuAspGluSerGluCysGluGly 327
QY 850 ATGAAGTTTGCACCAAAATGCCAAATGGTTGCGAGGATTTGATTTCCACCTCAACACAG 909
DB 328 Met-----LeuGluThrLeuProArg---CysLeuSerLeuIleGluSerCysTyrSer 344
QY 910 ACNAACCGCACCGCATTAGCTAGTCCCTCTGCGCGGAGCCACCAACATGTCAGG 969
DB 345 SerGln-----SerValPheSerCysValProAlaSerIleTyrCysAsn 359
QY 970 GACAATGTCAGGGGCCATACAGCTTTCGTGGTGGTGGTGTATGATATTCGGCAT 1029
DB 360 AsnAlaGlnLeuGlyProPheGlnLys---ThrGlyArgAsnValTyrAspValArgLys 378

```

```

QY 1030 CCA-----TATGATGACCCGACTCCGCCAAGTTATATACACAAATTT 1071
DB 379 MetCysGluGlyThrLeuCysTyrLysAsp-----MetGluTyrIleAspGlnTyr 395
QY 1072 CTGGCAAGAGGACTCTGTCATGACGCTATCGCGGTCAACATC---AATACACCCAGTCC 1128
DB 396 LeuAsnGlnAspPheValLysGluLysIleGlyAlaGluValGluThrTyrGluSerCys 415
QY 1129 AATAATGACGCTACTACGCTTCCAGCAAAACAGCGACTTGTCTGGCC----- 1179
DB 416 AsnPheAspValAsnArgAsnPheLeuPheAlaGlyAspIrpMetLysProTyrHisLys 435
QY 1180 AACITTCAGAGACTCGAGGAGATCTCTCTCCCGCGTGTCTCCCTCATCAT 1239
DB 436 AsnValIleAsnLeuLeuGluGln-----GlyLeuProVal-----LeuIleTyr 450
QY 1240 ---GGCGAGCGGATTACATCTGCAACTGGTTCGGCGGTGAGCGGTTTCCCTCGCTCGG 1296
DB 451 AlaGlyAspLysAspPheIleCysAsnIrpLeuGlyAsnGlnAlaIrpSerAsnGluLeu 470
QY 1297 AACTACTCCCAAGCGCGCTCGAGTTCGAAAGCGCAGGG---TACACGCCCTGAAAGTCAAC 1353
DB 471 ProIrpSerGlyHisGluGluPheGluSerAlaGluLeuTyrAsnLeuThrLeuLysAsp 490
QY 1354 GCGCTCGAGTATGGGAAACTCGCGAGTATGTAATTTCTCTCTACCTCGGCTCATGAG 1413
DB 491 GlyThrLysValGlyGluValLysAsnAlaGlyLysPheThrPheAlaArgMetPheAsp 510
QY 1414 GCAGGCCAAGTCCCATACTACAGCCCATCGCTCTCCCTCCCAATTTTAAACGGGACT 1473
DB 511 GlyGlyHisMetValProTyrAspGlnProGluSerLeuAlaMetValAsnArgIrp 530
QY 1474 ATCTTCGTTGGTATATCGCAGAGGCGCAGAGAAG 1509
DB 531 IleAlaGly---AspTyrSerLeuGlyThrLysLys 541

```

## RESULT 5

```

OL4414 ID OL4414 PRELIMINARY; PRT; 537 AA.
AC OL4414;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DI 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase Y.
GN CPY.
OS Pichia angusta (Yeast) (Hansenula polymorpha).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4905;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DL-1;
RA Bae J.-H., Kim H.-Y., Sohn J.-H., Choi E.-S., Rhee S.-K.;
RT "Molecular cloning and characterization of a gene coding for
RI carboxypeptidase Y from yeast Hansenula polymorpha DL-1."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U67174; AAB68520.2; -.
DR HSSP; P00729; 1CPY.
DR MEROPS; S10.001; -.
DR InterPro; IPR001563; Serine_carbpept.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR ProDom; PD001189; Serine_carbpept; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
SQ SEQUENCE 537 AA; 60270 MW; FC0LE57D7582C941 CRC64;

```

## Alignment Scores:

Pred. No.: 2,67e-34 Length: 537  
 Score: 595.50 Matches: 165  
 Percent Similarity: 45.01% Conservative: 65  
 Best Local Similarity: 32.29% Mismatches: 196

```

Query Match: 20.51% Indels: 85
DB: 3 Gaps: 20
us-09-712-338-1_copy_55_1662 (1-1608) x 014414 (1-537)

QY 37 CAGCTACCAAGAACCCACCGGGGTCAGAGCTCTTACAAACCGCAACAATGTACCAATC 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 92 LysileThrLysLysProAlaAspPheTrpAspPheAsnValLeuAsnGluLysPheSer 111
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 97 CGGTACAAG-----GACCGGGGCGAGGCGGCTCGAGACTAC 138
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 112 AsnTyrLysLeuArgValLysLysThrAspProGlyAlaLeuGlyLeuAspHisThr--- 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 139 CCGGGTCAAACTCTGATATGTCAGACCTCTCCCGAG---TCCCATACCTTC 195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 -----ArgGlnTyrSerGlyTyrLeuAspValGluAspGluAspLysHisPhe 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 196 TTCTGCTTCGCAAGCAGACATACCCAGAACTGCACCTATPCACATTGTGTTCAAT 255
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 148 TyrTrpMetPheGluSerArgAsnAspProValAsnAspProValLLeuLeuTrpLeuAsn 167
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 256 GGTGGCCCTGGAAGCGATTCTTTGATCGGTCTCTTCGAGAGTGGCCCTTGCCATGTC 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 168 GlyGlyProGlyCysSerSerLeuThrGlyMetLeuPheGluLeuGlySerAlaSerIle 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 316 AATTGCACTTTGATGACTACATCAACCTCACCTCGTGGAAACGAGGCTCCAAATTACTA 375
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188 GlyProAspLeuLysProIleAsnAsnProTyrSerTrpAsnSerAlaThrValle 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 376 TTCCTGTCCGAGCATTTGGAGTCGGCTTTTCATATAGTATGATAGCGTTGATGGTCCATT 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 PheLeuAspGlnProValAsnValGlyPheSerTyrSer----- 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 436 AACCTGTAAGTGGGTGCTGGAATTCGAGCTTTGCAGGAGTTCAGGGCGGTACCCA 495
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 ----- 220

QY 496 ACCATTGATGCACTCTGATGATACTACCAATCTGCCGAGAGCGGCTTGGGAGATC 555
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 221 -----SerLysSerValSerAsnThrValAlaAlaGlyLysAspValTyrAlaPhe 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 556 CTGAAGGATTCCTTAGTGGACTACCTAGCTGGAGTCTAGGGGCGAGCTCAAGAGCTTC 615
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 LeuGluLeuPheTyrGlnGlnPheProHisLeu-----LeuLysAsnAspPhe 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 616 AGTCTATGCGAGAGCTATGGAGGCGACTATGTCCTGCACTTCTCAATCAATTTCAC 675
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 HisIleAlaGlyGluSerTyrGlyGlyHisTyrIleProValPheAlaSerGluLeu 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 676 GAGCAGAATGAGAGAATGCCAACGGTAGTGTATGTTGCTAGCTTAATTCACACTCT 735
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 274 ThrHisAlaAspArg-----SerPheAsnLeuThrSer 284
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 736 CTGGGAATATTAAGCGCATATCGACAGGCGATCCAGGCGCTTACTACCTCAATTC 795
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 285 ValLeuIleGlyAsnGlyLeuThrAspProLeuAsnGlnTyrProPheTyrGluArgMet 304
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 796 GCTGTGACAATACCTACGAGTATCAAG--GCTGTCAACGACGAGCGTCTACAACTACATG 852
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 305 AlaCysSerThrAspGlyGlyTyrGluProThrIleAspGluSerGluCysGluGlyMet 324
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 853 AAGTTTGCAACCAATGCCAAATGGTTGCCAGGATTTGATTTCCACTGCCCTGCAACAGACA 912
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 325 -----LeuGluThrLeuProArg---CysLeuSerLeuIleGluSerCysTyrSerSer 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 913 AACCCACCGCATTAGCTAGCTACAGCCCTCTCCGCGGAGCCACCACAACTATGCGAGGAC 972
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 342 Gln-----SerValPheSerCysValProAlaSerIleTyrCysAsnAsn 356
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 973 AATGTTGAGGGGCCATACACGCTTTGCTGCTGCTGTTGTATGATATATCGCGATCCA 1032
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 AlaGlnLeuGlyProPheGlnLys---ThrGlyArgAsnValTyrAspValArgLysMet 375
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 6

```

Q8VWQ0 PRELIMINARY; PRT; 507 AA.
ID Q8VWQ0;
DC Q8VWQ0;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Putative serine carboxypeptidase.
GN SCP.
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX NCBI_TaxID=3635;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Xu-142; TISSUE=Fiber;
RA Li C.-H., Zhu Y.-Q., Meng Y.-L., Wang J.-W., Xu K.-X., Zhang T.-Z.,
RA Chen X.-Y.;
RT "Isolation of Genes Preferentially Expressed in Cotton Fiber Cells by
RT cDNA Array and RT-PCR.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY072822; AAL67992.1;
DR MEROPS; S10.009;
DR InterPro; IPR001563; Serine_carbpept.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR ProDom; PD001189; Serine_carbpept; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
DR Carboxypeptidase.
SQ SEQUENCE 507 AA; 56594 MW; 6A0D283D9C5E7173 CRC64;

```

## Alignment Scores:

```

Pred. No.: 6,06e-26 Length: 507
Score: 479.00 Matches: 145

```

Percent Similarity:	44.27%	Conservative:	83
Best Local Similarity:	28.16%	Mismatches:	19
Query Match:	16.49%	Indels:	96
DB:	10	Gaps:	20

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x Q8VWQ0 (1-507)

31	AGAAAGCAGCTACCCAAAGAACCCCAACGGGGTCAAGACTCTTACAAACGCAACCAATGTC	90	
QY			
Db	ArgValSerLeuProGluGlyProGlyLeuValGlnLys	77	
QY	91	ACCATCCGGTACAAAGAACCCGGGCAGAGCGGCTCTCGAGACTACCCGGGGTGTCAA	150
Db			
Db	ArgPheLysPheProThrLeuGlu	ValProGlyGlyVal	90
QY	151	TCCTAC	186
Db			
Db	SerPheGluAspLeuGlyHisHisAlaGlyTyrTyrLysLeuProAsnSerHisAspAla	110	
QY	187	CATACCTTCTCTGGTTCTTCGAAGCCAGACATAACCCAGAACTGCACCHTATCATATG	246
Db			
Db	ArgMetPheTyrPhePheGluSerArg	AsnSerLysLysAspProValValIle	129
QY	247	TGTTGAATGGTGGCCCTGGAAGCGATCTTTGATCGGTCTCTCGAAGACTTGGCCCT	306
Db			
Db	TrpLeuThrGlyGlyProGlyCysSerSerGluLeuAlaLeuPheTyrGluAsnGlyPro	149	
QY	307	TGCCAATGTCAAATCGACTTTTGTGACTACATCAACCTCTACTCGTGGACGAGTCTCC	366
Db			
Db	PheThrIleAlaAspAsnMetSerLeuValTrpAsnGluTyrGlyTrpAspLysAlaSer	169	
QY	367	AAATTACTATTCCTGTCTGCCACCATTCGGAGTCGGCTTTTCATATAGTATACGGTTGAT	426
Db			
Db	AsnLeuLeuTyrValAspGlnProIleGlyThrGlyPheSerTyrSer	185	
QY	427	GGGTCCATTAAACCTGTAACTGGGTCGTGCGAAATTCGAGCTTTCGAGGAGTTCAGGGC	486
Db			
Db	CGGTACCCCAACCAATGATGCCACTCTGATCGATACTACCAATCTTCGGCAGAGCCGCT	546	
QY	194	AsnGluAspGluValSerAsnAspLeuTyrAsp	204
Db			
QY	547	TGGGAGATCCTCGAAGATTCCTTAGTGGACTACTAGTCTAGTCTAGGTCAGTCT	606
Db			
Db	PheLeuGlnAlaPheAlaGluHisProGluPhe	AlaLys	218
QY	607	AAGGACTTCAGTCTATGGACGGAGCTATGGAGGCGCACTATGCTCTGCATCTTCAAT	666
Db			
Db	AsnAspPheIleThrGlyGluSerTyrAlaGlyHisTyrIleProAlaPheAlaAla	238	
QY	667	CATTTTACACGAGATCAGAGAAATGCGACGGTAGTGTAAATGGTGTCTAGCTTAAT	726
Db			
Db	ArgValHisArgLysAsnLys	AlaLysGluGlyIleHisIleAsn	253
QY	727	TTCAACTCTCTGGGAATTATTAAACGGCAATCATCGACGAGCGGATCCAGCCCTTACTAC	786
Db			
Db	LeuLysGlyPheAlaIleGlyAsnGlyLeuThrAspProAlaIleGlnTyrLysAlaTyr	273	
QY	787	CCTGAATTCGCTGAACATACCTACGGTATCAAGGCTGTCAACGAGACCGCTACAAC	846
Db			
Db	ThrAspTyrAlaLeuAspMetGlyValIleLysLysSerAspHisAspArgIle	291	
QY	847	TACATGAAGTTTGCCAAACCAATGCCAAATGGTTGCCAGGATTTGATTTCCACCTGCAAA	906
Db			
Db	AsnLysLeuValProValCysGluMetAlaIleLysLeuCysGly	306	
QY	907	CAGACAAACCGC	963
Db			
Db	ThrAspGlyThrIleSerCysMetAlaSerTyrPheValCys	320	
QY	964	TCCAGGGACAATGTCAGGGCCCACTACGCGCTTGCTGGT---	1020
Db			

```

321  -----AsnAsnIlePheAsnGlyIleMetAlaLeuAlaGlyAspThrAsnTyrTyrAsp 338
QY 1021 ATTCGGCATCAATGATGACCCGAGCTCCGCCAAGTATTACAAC-----AAATTTCGTG 1074
    ::::||||| ::: ::: :||| :|||||
Db 339 ValArgLysLysCysGluGlySerLeuCysTyrAspPheSerAsnMetGluSerPheLeu 358
QY 1075 GCAAGGACTCTGTCATGAGCGATCATGGCGTC---AACATCAACTCAACCCAGCTCCAAT 1131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| :::
Db 359 AsnLysLysSerValArgAspAlaLeuGlyValGlyAsnIleAspPheValSerCysSer 378
QY 1132 AATGACGCTACTACGCTTCCACCAACAGCGGCACTTTGCTGGCCCAACTCAATCGAA 1191
    ||||| ||||| ||||| ::: ::: :||| :|||
Db 379 ProThrValTyrGlnAla-----MetLeuValAspTyrMetArg 391
QY 1192 GACCTCAGGAGATCCTTGCTCTCCCGTGGTGTCTCC-----CTCATC 1236
    ::::||||| ::: ::::||||| :||| :||| :|||
Db 392 AsnLeuGlu-----ValGlyIleProValLeuLeuGluAspGlyIleLysLeuVal 409
QY 1237 TAT---GGCGAGCCGCAATTACATCTGCAACTGTTCTGGCGGTCAGGCGGTTCCCTCGCT 1293
    ||| ||||| ||| ||||| ||||| ||| |||
Db 410 TyrAlaGlyGluTyrAspLeuIleCysAsnTrpLeuclyAsnSerArgTTPValHisAla 429
QY 1294 GCGAACTACTCCCAAGCGCCGAGTTCGGAAGCGCAGGGTACAGCCGCCCTGAAGTCAAC 1353
    :||| :||| :||| :||| :||| :||| :||| :|||
Db 430 MetGluTrpSerGlyGlnLysGluPheValAlaSerProGluValProPheIleValAsp 449
QY 1354 GCGCTCAGTATGGGAAACTCGCGAGTATGGTAAATTTCTCTCTACTCCGCTCTATGAG 1413
    ||| ||| ||| ||| :||| :||| ||| ||| :||| :||| :|||
Db 450 GlyAlaGluAlaGlyValLeuLysThrHisGlyAlaLeuGlyPheLeuLysValHisAsp 469
QY 1414 CGAGGCCATGAAGTCCCATACTACTACGCCCATCGCTCCCTGCAATTTGTTTAACGCGACT 1473
    ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 470 AlaGlyHisMetValProMetAspGlnProLysAlaAlaLeuGluMetLeuLysArgTTP 489
QY 1474 ATCTTCGGTGG-----GATATCGCAGAGGGCCAGAGAAGATC 1512
    ||| ||| ||| :||| :||| :||| :|||
Db 490 ThrLysGlyThrLeuSerAspAlaSerAspSerGluLysLeuVal 504

RESULT 7
ID Q8GVTL PRELIMINARY; PRT; 429 AA.
AC Q8GVTL
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase C cbp31.
GN P043A03.7.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartolidae; Oryzeae; Oryza.
ON NCBI_TaxID=39947;
RX [1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT *Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC
RT clone:p043A03.7.
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RS EMBL; AP004299; BAC45113.1; -
SQ SEQUENCE 429 AA; 47790 MW; 1F2F64F2364756B4 CRC64;

```

Alignment Scores:

Pred. No.:	1.55e-25	Length:	429
Score:	473.00	Matches:	130
Percent Similarity:	45.4%	Conservative:	77
Best Local Similarity:	28.5%	Mismatches:	178
Query Match:	16.2%	Indels:	70
DB:	10	Gaps:	16

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x Q8GVT1 (1-429)

154 TACICTGGATAT-----GTCGACACCTCTCCCGAGTCCCATACCTTCTTCTGGTTCTTC 207

Db 17 HisAlaGlyTyrArgLeuProAsnThrHisAspAlaArgLeuPheTyrPhePhe 36  
QY 208 GAAGCCAGACATAACCA---GAACTGCACATATACATTTGGTTGAATGGTCCCT 264  
Db 37 GluSerArgGlySerGlyGluAspProValValIleTyrPheGlyPro 56  
QY 265 GGAAGCGATCTTGTATCGCTCTCTCGAAGAGTTGGCGCTTGCATGTCATTCGACT 324  
Db 57 GlyCysSerSerGluLeuAlaLeuPheTyrGluAsnGlyProPheHisIleAlaAsp 76  
QY 325 TTTGATGACTACATCAACCTCAGCTGCGAAGAGTCTCCAAATTTACTATTCTGTC 384  
Db 77 MetSerLeuValTyrAsnAspPheGlyTyrAspGlnGluSerAsnLeuIleTyrValAsp 96  
QY 385 CAGCCATTGGAGTCGGCTTTTCATATAGTATACGGTTGATGGGTCCATTACCCCTGTA 444  
Db 97 GlnProThrGlyThrGlyPheSerTyrSerAsnProArgAspThrArgHisAspGlu 116  
QY 445 ACTGGGTCGTCGAAATTCGAGCTTTGAGGAGTTTCAGGGCGGTACCCACCATGTAT 504  
Db 117 AlaGly--ValSerAsnAspLeuTyrAla----- 125  
QY 505 GGCATCTGATCGATACTACCAATCTTGGCGAGAGCGCTTGGAGATCTCGCAAGGA 564  
Db 126 -----PheLeuGlnAla 129  
QY 565 TTCCTAGTGGACTAGTCTGAGTCTAGGCTCAGTCTAAGACTTCAGTCTATG 624  
Db 130 PhePheThrGluHisProAsnPhe-----AlaLysAsnAspPheTyrIleThr 145  
QY 625 ACGGAGAGCTATGAGGGCACTATGCTCTGCATTTCTCAATCTTACGAGCAGAT 684  
Db 146 GlyGluSerTyrAlaGlyHisTyrIleProAlaPheAlaSerArgValTyrLysGlyAsn 165  
QY 685 GAGAGATTCACCAAGCTAGTGTAAATGGTTCAGCTTAATTTCAACTCTCGGGAAAT 744  
Db 166 Lys-----AsnSerGluGlyIleHisIleAsnLeuLysGlyPheAlaIle 180  
QY 745 ATTAACGGCATCTGAGGAGCGATCCAGGCCCTTACTACCTGATTCGCTGTAAC 804  
Db 181 GlyAsnGlyLeuThrAspProAlaIleGlyTyrLysAlaTyrThrAspTyrSerLeuLeu 200  
QY 805 AATACCTACGATATCAGGCTGTCAGGAGACCGTCTPACAACTACATGAATTTGCCAAC 864  
Db 201 -----MetGlyLeuIleThrLysSerGlnPheAsnArgIle-----Asn 213  
QY 865 CAATGCCAATGTTCCAGGATTTGATTTCCACCTGCAACACAGACAAACCGC---ACC 921  
Db 214 LysIleValProThrCysGluLeuAlaIleLysLeuCysGlyThrSerGlyThrIleSer 233  
QY 922 GCATTAGTCTAGTACGCCCTCTCGCGGAGACCCACACATGTGCAGGACAAATGTTGAG 981  
Db 234 CysLeuGlyAlaTyrValCys-----AsnLeuIlePheSerSerIleGlu 249  
QY 982 GGGCCACTACGCTTGTGTCGTGTGTATGATATTCGGCATCCA----- 1032  
Db 250 -----ThrIleGlyLysLysAsnTyrTyrAspIleArgLysProCysValGly 266  
QY 1033 -----TATGATACCGGACTCCCGCAAGTTATTACACAAATTTCTGGCAAGGAC 1083  
Db 267 SerLeuCysTyrAsp-----LeuSerAsnMetGluLysPheLeuGlnLeuLys 282  
QY 1084 TCTGTATGAGGCTATCGGCTC---AACATACTACACCCAGTCCAAATAAGACGCT 1140  
Db 283 SerValArgGluSerLeuGlyValGlyAspIleGlnPheValSerCysSerProThrVal 302  
QY 1141 TACTACGCTTTCCAGCAACAGCGGACTTTGTCTGGCCCACTTCAATCGAA---GACCTC 1197  
Db 303 TyrGlnAlaMetLeu-----LeuAspTyrMetArgAsnLeuGluValGlyIle 318  
QY 1198 GAGGAGATCCCTGCTCTCCCGTGGTGTCTCCCTCATCTATGCGGAGCGCGATTCATC 1257

Db 319 ProGluLeuLeuGluAsnAspIleLysValLeuIleTyrAlaGlyGluTyrAspLeuIle 338  
QY 1258 TGCAACTGTTGCGCGTCCAGGCGCTTCCCTCGCTCGGAACTPACCCAGCCGCCAG 1317  
Db 339 CysAsnTripleGlyAsnSerArgTyrValAsnSerMetGluTyrPheGlyLysGluAla 358  
QY 1318 TTCGGAAGCGGAGGTACAGCCCTCGAAGTCAACGGCTCGAGTATGGGAACTCGC 1377  
Db 359 PheValSerSerGluGluProPheThrValAspGlyLysGluAlaGlyIleLeuLys 378  
QY 1378 GAGTATGGTAATTTCTCTCCTGCTATGAGGAGCGCATCAAGTCCCATCTACTAC 1437  
Db 379 SerTyrGlyProLeuSerPheLeuLysValHisAspAlaGlyHisMetValProMetAsp 398  
QY 1438 CAGCCATCGCTCCCTGCAATTTTAAACCGGACTATCTCGGT 1482  
Db 399 GlnProLysValAlaLeuGluMetLeuMetArgTyrPheThrSerGly 413  
RESULT 8  
Q9FFB0 PRELIMINARY; PRT; 505 AA.  
ID AC Q9FFB0  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Serine carboxypeptidase.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
CX NCBI\_TaxID=3702;  
RN NCBI\_Locus=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RX MEDLINE=97471969; PubMed=9330910;  
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
RA Miyajima N., Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
RT features of the 1.6 Mb regions covered by twenty physically assigned  
RT pl clones."  
RL DNA Res. 4:215-230(1997).  
DR EMBL; AB005243; BAB10619.1; -  
DR HSPF; P10619.1; IVY.  
DR MEROPS; S10.009; -  
DR InterPro; IPR001563; Serine\_carbpept.  
DR InterPro; IPR000379; Ser\_estrs\_site.  
DR Pfam; PF00450; serine\_carbpept; 1.  
DR PRINTS; PR00724; CRBOXYPTASEC.  
DR ProDom; PD001189; Serine\_carbpept; 1.  
DR PROSITE; PS00560; CARBOXYPEPT\_SER\_HIS; 1.  
DR PROSITE; PS00131; CARBOXYPEPT\_SER\_SER; 1.  
KW Carboxypeptidase.  
SQ SEQUENCE 505 AA; 56543 MW; 6E74CD351090B099 CRC64;  
Alignment Scores:  
Pred. No.: 4,79e-25 Length: 505  
Score: 466.50 Matches: 132  
Percent Similarity: 44.42% Conservative: 75  
Best Local Similarity: 28.33% Mismatches: 178  
Query Match: 16.06% Indels: 81  
DB: 15 Gaps:  
US-09-712-338-1\_COPY\_55\_1662 (1-1608) x Q9FFB0 (1-505)  
QY 139 CCGGTGTGCAATCC-----TACTCTGGATAT-----GTGACACCTCTCCCGAG 183  
Db 91 ProSerValGlnGluPheGlyHisTyrAlaGlyTyrSerLeuProHisSerLysSer 110  
QY 184 TCCCATACCTTCTTCTGTTCTTCCAGCCAGACATACCCAGAACTGCACCTATACA 243  
Db 111 AlaLysMetPheTyrPhePheGluSerArg---AsnLysThrThrAspProValVal 129  
QY 244 TTGTGGTTGAATGGTGGCGCTCGAAGCGATTCTTTGATCGGTCTCTTCGAGAGTTGGGC 303



```

Db 130 ILETRPLEUHRGLYGLYPROGLYCYSERVALALAMETPHETYRGLUASNDGLY 149
QY 304 CTTGCCATGCAATTCGACTTTTGAUGACTACATCAACCTCTACTCGTGAACAGAGTC 363
Db 150 ProPheLysLeuSerLysAspLeuSerLeuTyrTrpAsnAspPheGlyTrpAspLysVal 169
QY 364 TCCAAATTTACTATTCCTGTCCACCCATPTGGAGTCGCGCTTTTCATATAGTAGTACGGTT 423
Db 170 SerAsnLleLeuValAspGlnProValGlyThrGlyPheSerTyrThrSerAspGlu 189
QY 424 GATGGTCCATTAACCTGTAACTAGGGGTCTCCAAAATTTCGAGCTTTCGAGGAGTTCAG 483
Db 190 SerAspLleArgAsnAspGluAspGlyVal----- 199
QY 484 GGCGGTACCAACCAATGTATGTCACCTCTGATCATCTACCATCTCTGCGCGAGAGGCC 543
Db 200 -----SerAsnAsp 202
QY 544 GCTTGGAGATCCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGTCGAG 603
Db 203 LeuTyrAspPheLeuGlnAlaPhePheLysGluHisPro-----LysPheVal 218
QY 604 TCTAAGGACTTCAGTCTATGAGCGAGGAGCTATGAGGCGACTATGCTCCTGCTATCTTC 663
Db 219 LysAsnAspPhePheLleThrGlyGluSerTyrAlaGlyHisTyrLleProAlaLeuAla 238
QY 664 AATCATTTTACGAGCAGATGAGAGAAATGCCAAACGGTAGTGTATGATGTGTTCAGCTT 723
Db 239 SerArgValHisSerGlyAsnLysLys-----LysGluGlyIleProIle 253
QY 724 AATTCACATCTCTGGAAATTAATAAGGCATCATCGACGAGCGATCCAGGCGCCCTTAC 783
Db 254 AsnLeuLysGlyPheAlaLleGlyAsnGlyLeuThrAsnProGluLleGlnTyrGlyAla 273
QY 784 TACCCGTGAATTCCTGTGAACAATACCTACGGTATCAAGCGCTGTCACGAGACCGTCTAC 843
Db 274 TyrGlyAspTyrAlaLeuGln-----MetLysLeuLleSerGluSerAspHis 289
QY 844 AACTACATCAAGTTTCCCAACCAAAATGCCAAATGGTTCGACGATTTGATTTCCACTGC 903
Db 290 GluSerLysLys-----GlnAspTyrValGluCysGln 300
QY 904 AAACAGACAACCGCACCGCATPAGCTGACTAGCCCTCTGCGCCGAGCCACCAACATG 963
Db 301 AsnLleThrLysLysCysSerLeuGlyGlyGlyLeuValCysAspSerAlaValGluVal 320
QY 964 TGCAGGACAATGTTTCAGGGCCATACACTAGCCTTTTGGTGGTGTGTGTATGATATT 1023
Db 321 CysThrSerLlePheAsnLysLleValAlaLysLysSerGlyLeuAsnTyrTyrAspIle 340
QY 1024 CGGCATCCA-----TATGATGACCCGACCTCCGCGCAAGTATTATACAAC 1065
Db 341 ArgLysLysCysValGlySerLeuCysTyrAspPhe-----SerArgMetGlu 356
QY 1066 AAATTTCTGGCAAGACTCTCTCATGAGCGATATCGGGTC---AACATCAACTACACC 1122
Db 357 IlePheLeuAsnLysGluAsnValArgLysAlaLeuGlyValGlyAspLleLysPheVal 376
QY 1123 CAGTCCAAATAGCAGCTACTACGCTTTCACGAAACAGGCGACTTCTCTGCGCCAC 1182
Db 377 SerCysSerSerThrValTyr-----AspAlaMetIleGluAsp 389
QY 1183 TTTCATCGAAGACTCGAGGATGCTTGTCTCTCCCGTTCGCTGCTCC----- 1230
Db 390 TrpMetGlnAsnLeuGlu-----ValLysIleProSerLeuValAsnAspGlyIleAsn 407
QY 1231 ---CTCATCTAT---GGCGACCGCGATTCATCTGCAACTGTTTCGGGGTTCAGCGGTT 1284
Db 408 LeuLeuValTyrAlaGlyCyluTyrAspLleLleCysAsnTyrPleuGlyAsnSerArgTrp 427
QY 1285 TCCCTCGCTGCAACTACTCCCAAGCGCCCGCTTCGGAAGCGGAGGTACACGCCCTG 1344

```

```

Db 428 ValAspGlnMetAsnTrpSerGlyGlnLysGlyPheGlySerAlaLysAsnValSerPhe 447
QY 1345 AAGTCAACGCGTCGAGTATGGGAAATCGCGAGTATGTAATTTCTCCCTCACTGCGC 1404
Db 448 LeuValAspGlyLysGluAlaGlyLeuLeuLysAsnHisGlyProLeuThrPheLeuLys 467
QY 1405 GTCTATGAGGAGCGCATGCAAGTCCCATCTACTACAGCCCATCGCTCCCTGCAATGTTT 1464
Db 468 ValTyrAsnAlaGlyHisMetValProMetAspGlnProLysAlaSerLeuGlnMetLeu 487
QY 1465 AACCGGACTATCTTCGGT 1482
Db 488 GlnAsnTrpMetGlnGly 493

RESULT 9
Q9XH61 PRELIMINARY; PRT; 501 AA.
AC Q9XH61;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-MAR-2002 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
OS Serine carboxypeptidase.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC Anthemideae; Matricaria.
OX NCBI_TaxID=98504;
RN [1]
RP SEQUENCE FROM N.A.
RA Kohchi C., Yasuda H., Hirata T.;
RT "Isolation of a cDNA encoding for a carboxypeptidase, having leucine
RT zipper structure at the N-terminal region, from the cultured shoot
RT primordia of Matricaria chamomilla."
RL Plant Biotechnol. 16:409-412(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Kohchi C., Yasuda H., Hirata T.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF141384; AAD42963.2; -.
DR HSSP; P00729; LYSC.
DR MEROPS; S10.009; -.
DR InterPro; IPR001563; Serine_carbpept.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR PRODOM; PD001189; Serine_carbpept; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
KW Carboxypeptidase.
SQ SEQUENCE 501 AA; 55973 MW; DBFD8DB41860A66 CRC64;

Alignment Scores:
Pred. No.: 1-18e-24 Length: 501
Score: 461.00 Matches: 131
Percent Similarity: 44.33% Conservative: 76
Best Local Similarity: 28.05% Mismatches: 188
Query Match: 15.87% Indels: 72
DB: 10 Gaps: 15

US-09-712-338-1_COPY_55_1662 (1-1608) x Q9XH61 (1-501)
QY 115 GCAGAGGCGTCTGCGAGACTACCCGCGGTGCAATCTACTCTGATAT-----GTC 168
Db 80 SerAspGlyVal-----ThrValGluAspLeuGlyHisHisAlaGlyTyrTyrGlnIle 97
QY 169 GACACCTCTCCGAGTCCCAIACCTTCTTCTGTTCTTCGAGCCAGACATACCCAGAA 228
Db 98 GlnHisSerHisAlaLysMetPheTyrPhePheGluSerArgAsnAsnLysLys 117
QY 229 ACTGCACCATCATCATGTTGTTGAATGTTGGCCCTCGAAGCGATTCCTTGTATCGCTC 288
Db 118 Asp---ProValIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 136

```





Score: 439.00 Matches: 130  
Percent Similarity: 42.43% Conservatives: 83  
Best local Similarity: 25.90% Mismatches: 191  
Query Match: 15.12% Indels: 98  
DB: 10 Gaps: 16

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x Q9LXH4 (1-510)

QY 16 CCGGGCGTGGTGGTAGAGACAGCTACCAAGAACCCCGGGGTCAAGACTCTTACA 75  
|||||:|||||  
Db 82 ProAlaAlaValAspArg 88  
QY 76 ACCGCAACAATGTCAACCATCCGTACAGGAACCCGGGGCAGAGGGCGTCTGCGAGACT 135  
|||||:|||||  
Db 89 -----GlySerGlyGly----- 92  
QY 136 ACCCGGGGGTCAAAATCCTAC-----TCTGATAT-----GTCGACACCTCTCC 180  
|||||:|||||  
Db 93 SerProSerValGlnAspPheGlyHisAlaGlyTyrTyrLysLeuProAsnSerLys 112  
QY 181 GAGTCCCATACCTCTCTCGGTCTCTCGAGCCAGACATAACCCGAAACTGCACCTATC 240  
|||||:|||||  
Db 113 AlaAlaArgMetPheTyrPhePheGluSerArgThrAsn---LysAlaAspProVal 131  
QY 241 ACATTGTGTTGAATGGTGGCCCTGGAAGCGATTCTTTGATCGGTCTCTTCCGAAGAGTTG 300  
|||||:|||||  
Db 132 ValIleTyrPheThrGlyGlyProGlyCysSerSerGluLeuAlaLeuPheTyrGluAsn 151  
QY 301 GGCCCTGGCATGTCATTCGACTTTTGATGACTATCATCAACCTCCTCGTGAACGAG 360  
|||||:|||||  
Db 152 GlyProPheThrValSerAsnAsnSerSerLeuSerTyrAsnGluPheGlyTyrAspLys 171  
QY 361 GTCCTCAATTTACTATTCCTGTCCTCCAGCATGGGTGGTCTTTTCATAGTATGATACG 420  
|||||:|||||  
Db 172 AlaSerAsnLeuIleTyrValAspGlnProValGlyThrGlyPheSerTyrThrSerAsp 191  
QY 421 GTTGATGGTTCCTAATCACTGCTGAGTGGGTGCTGGAATTCGAGCTTTGCGAGGAGTT 480  
|||||:|||||  
Db 192 GlnSerAspLeuArgHisAspGluAspGlyVal----- 202  
QY 481 CAGGGCGGTACCCCAACCATTTGATGCCACTCTGATCGATFACCTACCAATCTGCGGAGAG 540  
|||||:|||||  
Db 203 -----SerAsn 204  
QY 541 GCCCTTGGGAGATCCTCGAAGATTCCTTAGTGAGTACCTAGCTTGGACTCTAGAGTG 600  
|||||:|||||  
Db 205 AspLeuTyrAspPheLeuGlnAlaPhePheLysGluHisPro-----GlnPhe 220  
QY 601 CAGTCTAAGGACTTCACTGATGACGAGGAGTATGAGGCGCACTATGTCCTGCAATC 660  
|||||:|||||  
Db 221 VallysAsnAspPheTyrIleThrGlyGluSerTyrAlaGlyHisTyrIleProAlaLeu 240  
QY 661 TTCATCATTTTACGAGCAGATGACGAGATTCCTCAACGGTACTGTTAATGGTGTTCAG 720  
|||||:|||||  
Db 241 AlaSerArgValHisArgGlyAsnLys-----AsnLysGluGlyThrHis 255  
QY 721 CTTAATTTCACTCTCTCGGAATTTAAGCGCATCATCGAGGAGGCGGATCCAGGCCCT 780  
|||||:|||||  
Db 256 IleAsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAsnProGluIleGlyTyrGly 275  
QY 781 TACTACCTCATTCGCTGTGAACAATACCTACGATATCAAGGCTGTCAACGACGCGTC 840  
|||||:|||||  
Db 276 AlaTyrAlaAspTyrAlaLeuAspMetAsnLeuIleThrGlnSerAspHisAspAsnLeu 295  
QY 841 TACAACTACATGACGTTGGCAACCAATGCCAATGTTGCCAGGATTTGATTTCCACC 900  
|||||:|||||  
Db 296 AsnArgTyr-----TyrAlaThr 301  
QY 901 TGCACCAACAGAACCCGACCGCATTTAGCTGACTACGCC-----CTCTGCGCCGAA 951  
|||||:|||||  
Db 302 CysGlnGlnSerIleLysGluCysSerAlaAspGlyGlyGluGlyAspAlaCysAlaSer 321  
QY 952 GCCACCAACATGTGCGAGGACAAATGTTGAGGGGCCATACTACGCTTTGCTGGTCTG--- 1008

Db 322 SerTyrThrValCys---AsnAsnIlePheGlnLysIleMetAspIleAlaGlyAsnVal 340  
|||||:|||||  
QY 1009 GGTGTGTATGATATCGGCATCATATGATGACCGGACTCCGCCAAGITATTACAAAC--- 1065  
|||||:|||||  
Db 341 AsnTyrTyrAspValArgLysGlnCysGluGlySerLeuCysTyrAspPheSerAsnMet 360  
QY 1066 ---AAATTTCTGGCAAGGACTCTGTCATGAGGCGTATCGGGTGC---AACATCACTAC 1119  
|||||:|||||  
Db 361 GluAsnPheLeuAsnGlnLysSerValArgLysAlaLeuGlyValGlyAspIleGluPhe 380  
QY 1120 ACCAGTCAATAATGACGCTACTACGCTTTCACGCAAAACAGGCGACTTGTCTGGCCC 1179  
|||||:|||||  
Db 381 ValSerCysSerThrAlaValTyrGluAlaMetGlnMetAsp-----TrpMet 396  
QY 1180 AACTTTCATCGAA---GACCTCGAGGAGATCCTCTCTCCCTCGGTGCTCTCCCTCATC 1236  
|||||:|||||  
Db 397 ArgAsnLeuGluValGlyIleProAlaLeuGlnAspGlyIleLysLeuValTyr 416  
QY 1237 TATGGGAGCGGATTCATCTGCAACTGCTGCGGCGTCCAGGCGTCTTCCCTCGTGGC 1296  
|||||:|||||  
Db 417 AlaGlyGluTyrAspLeuIleCysAsnTrpLeuGlyAsnSerLysTrpValHisGluMet 436  
QY 1297 AACTACTCCCAACCGCCGAGTTCGGAAGCGGAGGTACACGCGCTGAAAGTCAACGCGC 1356  
|||||:|||||  
Db 437 GluTyrSerGlyGlnLysGluPheValAlaAlaAlaThrValProPheHisValAspAsn 456  
QY 1357 GTCGAGATATGGGAAATCTCGGAGTATGATGTAATTCCTCTACTCGCGTCTATGAGGCA 1416  
|||||:|||||  
Db 457 LysGluAlaGlyLeuMetLysAsnTyrGlySerLeuThrPheLeuLysValHisAspAla 476  
QY 1417 GGCGATGAATCCATCTACTACCGCCATCGCTCCCGCAATGTGTTAACCGGAGCATC 1476  
|||||:|||||  
Db 477 GlyHisMetValProMetAspGlnProLysAlaAlaLeuGlnMetLeuGlnAsnTrpMet 496  
QY 1477 TTCGGT 1482  
Db 497 GlnGly 498

RESULT 12  
Q93ZC3 PRELIMINARY; PRT; 510 AA.  
ID Q93ZC3  
AC Q93ZC3; 2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE AT3g45010/F14D17.80.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
RA Kawai J., Lam J.M., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
RA Ecker J.R.;  
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AY057639; AAL15270.1;  
DR InterPro; IPR001563; Serine\_carbpept.  
DR InterPro; IPR000379; Ser\_estrs\_site.  
DR Pfam; PF00450; serine\_carbpept; 1.  
DR PRINTS; PR00724; CRBOXYP1ASEC.  
DR PRODOM; PD001189; Serine\_carbpept; 1.  
DR PROSITE; PS00560; CARBOXYPEPT\_SER\_HIS; 1.  
DR PROSITE; PS00131; CARBOXYPEPT\_SER\_SER; 1.  
SQ SEQUENCE 510 AA; 56837 MW; 704BC00C4E64BF08 CRC64;

## Alignment Scores:

Pred. No.: 5,33e-23 Length: 510  
 Score: 438.00 Matches: 130  
 Percent Similarity: 42.23% Conservative: 82  
 Best Local Similarity: 25.90% Mismatches: 192  
 Query Match: 15.08% Indels: 98  
 DB: 1.0 Gaps: 16

US-09-712-338-1\_copy\_55\_1662 (1-1608) x Q93ZC3 (1-510)

QY 16 CCGCGTCGTCGCTGAGAGACACTACCCAAAGAACCCACCAGGGGTCAAGACICTTTACA 75  
 DB 82 ProalaAlaValAspArgArg  
 QY 76 ACCGCAACAATGTCCACATCCGCTACAGCAAGCAACCGGGGCGAGAGGGCGTCTGGCAGACT 135  
 DB 89 -----GlySerGlyGly----- 92  
 QY 136 ACCCGGGGTGCAATCTCTAC-----TCTGGATAT-----GTGACACCTCTGCC 180  
 DB 93 SerProSerValGlnAspPheGlyHisHisAlaGlyTyrTyrLysLeuProAsnSerLys 112  
 QY 181 GAGTCCCATACCTCTCTCTGCTTCTCGAAGCCAGACATAACCCAGAACTGCACCTATC 240  
 DB 113 AlaAlaArgMetPheTyrPhePheGluSerArgThrAsn---LysAlaAspProVal 131  
 QY 241 ACATGTGCTGTAATGTGGCCCTCGGAAGCAATCTTTGATCGGTCTCTCTCGAGAGTTG 300  
 DB 132 VallIleTrpLeuThrGlyGlyProGlyGlySerSerGluLeuAlaLeuPheTyrGluAsn 151  
 QY 301 GGCCTTGCCTGCAATCTGACATTTTGATGACTACATCAACCTCCTCACTCGTGGAAACGAG 360  
 DB 152 GlyProPheTrpValSerAsnAsnSerSerLeuSerTrpAsnGluPheGlyTyrAspLys 171  
 QY 361 GTCCTCAATTTACTTCTCTGCTCCAGCAATGGAGTCGGCTTTTCATATAGTGATACG 420  
 DB 172 AlaSerAsnLeuIleTyrValAspGlnProValGlyThrGlyPheSerTyrThrSerAsp 191  
 QY 421 GTTGATGGTCCATTAACCTGTAACTGGGTCTGCTGCAAAATTCGAGCTTGGAGAGTT 480  
 DB 192 GlnSerAspLeuArgHisAspGluAspGlyVal----- 202  
 QY 481 CAGGGCCGGTACCCCAACCAATGTATGCCACTCTGTGATGATGATGATGATGATGATGAT 540  
 DB 203 -----SerAsn 204  
 QY 541 GCCCTTGGGAGATCTGCAAGGATTCCTTAGTGGACTACTAGTGGACTCTAGGAGTG 600  
 DB 205 AspLeuTyrAspPheLeuGlnAlaPhePheLysGluHisPro-----GlnPhe 220  
 QY 601 CAGCTAAGCACTCAGCTAGTACGAGGAGCTGATGGAGGCACTATGGTCTGCTTC 660  
 DB 221 VallYsnAspPheTyrIleThrGlyGluSerTyrAlaGlyHisTyrIleProAlaLeu 240  
 QY 661 TTCAATCATTTTACGAGCAGATGAGAGAAATGCCAACGGTACTGTAAATGGTGTTCAG 720  
 DB 241 AlaSerArgValHisArgGlyAsnLys-----AsnLysGluGlyThrHis 255  
 QY 721 CTTAATTTCAACTCTCTGGGAATATTAAAGGCATCATGAGGAGGATCCAGGCCCT 780  
 DB 256 IleAsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAsnProGluIleGluTyrGly 275  
 QY 781 TACTACCTCAATTCGCTGTCACAAATCACTACGGTATCAAGGCTGTCAAGCAGACCTC 840  
 DB 276 AlaTyrAlaAspTyrAlaLeuAspMetAsnLeuIleThrGlnSerAspHisAspAsnLeu 295  
 QY 841 TACAACCTACATGAAGTTTGCCCAACCAATGCCAAATGTTGCCAGGATTGATTTCCACC 900  
 DB 296 AsnArgTyr-----TyrAlaThr 301  
 QY 901 TGCAACAGACAAACCCAGCCGCTATAGCTGACTACGCC-----CTCTGCGCGGAA 951  
 DB 951 ----- 951

DB 302 CysGlnGlnSerIleLysGluCysSerAlaAspGlyGlyGluGlyAspAlaCysAlaSer 321  
 QY 952 GCCACCAACATGTGCAGGACAAATTTGAGGGGCCATACCTACGCCCTTGTCTGCTGCT--- 1008  
 DB 322 SerTyrThrValCys---AsnAsnIlePheGlnLysIleMetAspIleAlaGlyAsnVal 340  
 QY 1009 GGTGTGATGATATTCCGGCATCCATATGATGACCCGACTCCGCCAGTTATTACAC--- 1065  
 DB 341 AsnTyrTyrAspValArgLysGlnCysGluGlySerLeuCysTyrAspPheSerAsnMet 360  
 QY 1066 ---AAATTTCTGGCAAAGGACTCTCATGTCAGCTGCGGTC---AACATCAACATAC 1119  
 DB 361 GluAsnPheLeuAsnGlnLysSerValArgLysAlaLeuGlyValIleAspIleGluPhe 380  
 QY 1120 ACCGATCCATATGACGTCTACTACGCTTTCACGCAAAACAGGCGACTTTGTCTGGGCC 1179  
 DB 381 ValSerCysSerThrAlaValTyrGluAlaMetGlnMetAsp-----TrpMet 396  
 QY 1180 AACITCATCGAA---GACCTCGAGGAGATCTCTGCTCTCCCGTGGTGTCTCCCTCATC 1236  
 DB 397 ArgAsnLeuGluValGlyIleProAlaLeuLeuGlnAspGlyIleLysLeuValTyr 416  
 QY 1237 TATGCGCAGCGGATTCATCATCTGCACTGGTTCGGCGGTTCAGGCGGTTTCCCTCGCTCG 1296  
 DB 417 AlaGlyGluTyrAspLeuIleCysAsnTrpLeuGlyAsnSerLysTrpValHisGluMet 436  
 QY 1297 AACCTACTCCAAAGCGCCCGTTCGGAAGCGGAGGTACAGCCCTGAAAGTCAACAGCG 1356  
 DB 437 GluTrpSerGlyGlnLysGluPheValAlaAlaThrValProPheHisValGlyAsn 456  
 QY 1357 GTCCAGTATGGGAAACCTCGGAGTATGATGATGATGATGATGATGATGATGATGATGAT 1416  
 DB 457 LysGluAlaGlyLeuMetLysAsnTyrGlySerLeuThrPheLeuLysValHisAspAla 476  
 QY 1417 GCCAATGAAGTCCCATACATACGACCCATCGCTCCCTGCAATTTTAAACGGGACTATC 1476  
 DB 477 GlyHisMetValProMetAspGlnProLysAlaAlaLeuGlnMetLeuGlnAsnTrpMet 496  
 QY 1477 TTCGGT 1482  
 DB 497 GlnGly 498  
 RESULT 13  
 Q17679 PRELIMINARY; PRT: 2105 AA.  
 AC Q17679;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Y16B4A.2 protein.  
 GN Y16B4A.2  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_taxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Thomas K.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y20203; CA94110.1; -.  
 DR EMBL; AL023825; CA94110.1; JOINED.  
 DR EMBL; Y93339; CA94110.1; JOINED.  
 DR EMBL; Y93339; CAB07544.1; -.  
 DR EMBL; AL023825; CAB07544.1; JOINED.  
 DR EMBL; Y20203; CAB07544.1; JOINED.  
 DR EMBL; AL023825; CA94110.1; -.  
 DR EMBL; Y93339; CA94110.1; JOINED.  
 DR EMBL; Y20203; CA94110.1; JOINED.  
 DR HSSP; F10619; LIVY.  
 DR WormPep; Y16B4A.2; CE21374.  
 DR InterPro; IPR001563; Serine\_carbpept.  
 DR InterPro; IPR000379; Ser\_estrs\_site.  
 DR Pfam; PF00450; serine\_carbpept; 4.

DR PRINTS; PR00724; CRBOXYPTASEC.  
 DR PRODOM; PD001189; Serine\_carbpept; 6.  
 DR PROSITE; PS00560; CARBOXYPEPT\_SER\_HIS; 4.  
 DR PROSITE; PS00131; CARBOXYPEPT\_SER\_SER; 2.  
 SQ SEQUENCE 2105 AA; 234335 MW; 139CB6E96D75B3FB CRC64;

## Alignment Scores:

Pred. No.: 4,92e-21 Length: 2105  
 Score: 413.50 Matches: 150  
 Percent Similarity: 38.52% Conservative: 85  
 Best Local Similarity: 24.59% Mismatches: 202  
 Query Match: 14.24% Indels: 173  
 DB: 5 Gaps: 26

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x Q17679 (1-2105)

QY 10 AGTACACGGGTCGTCGGTAGAGACAGTACCCAGAACCCACGGGGTCAAGACT 69  
 |||||  
 Db 1521 SerThrProAlaGlyIle-----AspValThrProGlnProLeuThrAsn 1535  
 QY 70 CTTACAACCGCAACAATGTCAACCATCCGGTACAGGAACCCGGCGAGAGGGCGTCTGC 129  
 |||||  
 Db 1536 ValIleSerGlnThrAsnCysThr-----ThrGlyGlnThrAspArgIle 1550  
 QY 130 GAGACTACCCCGGGTGTC-----AAATCCTACTCTCGTATATCTCGAC 171  
 |||||  
 Db 1551 IleAsnLeuProGlyLeuProAlaAspMetLeuPheLysGlnTyrSerGlyPheLeuAsp 1570  
 QY 172 ACCTCTCCGAGTCCCAATACCTCTTCTGTTCTTCGAAGCCACACATAACCCAGAACT 231  
 |||||  
 Db 1571 GlyLeuSerGlyHisLysValHisTyrTrpLeuValGluSerGluAsnAsnProSerThr 1590  
 QY 232 GCACCTATCATGTTGGTGAATGGTGGCCCTGGAAGCGATCTCTTGATCGGTCTCTTC 291  
 |||||  
 Db 1591 AspProLeuLeuTrpLeuAsnGlyGlyProGlySerSerLeuMetGlyLeuPhe 1610  
 QY 292 GAAGAGTGGGCCCTGCCATGTCAATTCGACTTTTGATGACTACATC---AACCCCTCAC 348  
 |||||  
 Db 1611 GluGluAsnGlyProPheArgValSerLysAspSerGlnThrLeuSerArgAsnProTyr 1630  
 QY 349 TCGTGGACGAGGCTCCAAATTTACTATTCTGTGCCAGCCATTTGGAGTGGCTTTTCA 408  
 |||||  
 Db 1631 SerTrpAsnLysPheAlaAsnValLeuTyrLeuGluSerProIleGlyValGlyTyrSer 1650  
 QY 409 TATAGTATACGGTGTGATGGTGTCCATTAACCTCTAAGTGGGTCTGCGAAATTCGAGC 468  
 |||||  
 Db 1651 Tyrala----- 1652  
 QY 469 TTTGACGAGTTCAGGCGCGGTACCCAAACCATGATGCCACTCTGATCGATPACTACCAAT 528  
 |||||  
 Db 1653 -----TyrAsnAsnThrAsnIleGlnTyrAspValThr--- 1664  
 QY 529 CTTGCGCAGAGCGCGCTGGGAGATCTCTGAAGGATTCCTTAGTGGACTACCTAGCTTG 588  
 |||||  
 Db 1665 ---ThrAlaGlnGluAsnTyrAlaAlaLeuLysSerPhePheAlaGlnTyrPro----- 1681  
 QY 589 GACTCTAGGTGTCAGTCTAAGCACTTCACTATGACGAGAGATGAGGAGCACAT 648  
 |||||  
 Db 1682 -----GlnTyrThrThrSerAspPheTyrThrThrGlyGluSerTyrAlaGlyValTyr 1699  
 QY 649 GGTCTCGATCTTCATCATTTTACAGCAGAGATGAGAAATGCCAAGGTAGTGT 708  
 |||||  
 Db 1700 LeuProGlyLeu-----SerAlaLeuValGlnGlyIleLys 1712  
 QY 709 AATGGT---GPTCAGCTTAATTTCACTCTCTGGGAATTTATACGGCATCATCGACGAG 765  
 |||||  
 Db 1713 SerGlyAspIleAsnIleAsnTyrLysGlyValSerIleGlyAsnGlyValIleAspLys 1732  
 QY 766 -----GCGATCCAGCGCCCTTACTACCTCGAATTCGCTGTAACAAT 807  
 |||||  
 Db 1733 ArgThrAspMetAsnSerGlnLeuHisTyrGlnTyr----- 1745  
 QY 808 ACCTACGGTATCAGGCTGTCAACGAGACCGCTCTACAAC----- 846

Db 1746 ---HisGly-----GlyIleSerAlaThrThrTyrAsnThrAlaLeuGlnLeuCysCys 1762  
 QY 847 -----TACATGAAGTCTTGGCAACCAATG 870  
 Db 1763 SerGlyAspGluPheLysCysArgPheSerAspArgMetThrAsnPheAsnSerIle 1782  
 QY 871 CCAATGGT-----TGCAGGAT-----TTGATTTCACCTGCAACACAGACA 912  
 |||||  
 Db 1783 ProTrpGlyAspLeuSerAspProCysTyrAspPheValValAlaThrGlyAlaGlnLeu 1802  
 QY 913 AACCGCACCGCATAGCTAGCTAGCCCTCTGCGCCGGAAGCCACCAACATGTGAGGAC 972  
 |||||  
 Db 1803 LeuLeuThrAlaPheAspProTyrAsnMetTyrGlnGlnCysTrpThrIle----- 1819  
 QY 973 AATGTTGAGGGCCATACACTACGCCCTTCTGCTGGTGTGTATGATATTCGGCATCCA 1032  
 |||||  
 Db 1820 -----Pro 1820  
 QY 1033 TATGATGACCGACTCCGCCAAGTTATTACACAAA----- 1068  
 |||||  
 Db 1821 TyrAsnAspThrThrProArgThrProTyrGlyGluThrTrpThrGlyLeuAsnTyrGlu 1840  
 QY 1069 -----TITCTG 1074  
 Db 1841 SerSerAspPheLeuAsnGlyTyrProCysTyrAspSerAlaMetGluGlyTyrLeu 1860  
 QY 1075 GCAAGGACTCTGTCTGATGAGCTATCGCGTC-----AACATCACTACACCCAGTCC 1128  
 |||||  
 Db 1861 AsnArgProValValArgLysAlaLeuAsnIleProAspSerValProTyrTrpAlaAla 1880  
 QY 1129 AATAATGACGCTACTAGCTTTCAGCAACAGCGGACTTTGTCTGCCCAACTTCATC 1188  
 |||||  
 Db 1881 AsnAsnAsnIleIleAsnAlaTyrAsnGlnGlnValAspSerIleValProAsn----- 1898  
 QY 1189 GAAGACTCGGAGACTCTTCTGCT-----CTCCCGCGCTGCTCCTCTCATCTAT 1239  
 |||||  
 Db 1899 -----LeuGlnIleIleMetAlaAsnAlaProAlaAsnPheLysMetLeuTyrSer 1916  
 QY 1240 GCGCAGCGGATATCATCTGCAACHTGCTGCGCGTCAG----- 1278  
 |||||  
 Db 1917 GlyAlaAlaAspThrMetValAsnTrpLeuGlyAlaGluIlePheThrAlaAsnAsnPhe 1936  
 QY 1279 ---GCGCTTCTCCTCGCTGCGAATCTCCCAAGCGCC-----CAGTCCGAAGC 1326  
 |||||  
 Db 1937 AlaAlaLeuGlyLeuThrThrSerSerProArgAlaGlnTrpThrPheGlnTyrAsnSer 1956  
 QY 1327 GCAGGTACACGCCCTGAAAGTCAACGGCTGAGTATGGGAAACTCGCGAGTATGCT 1386  
 |||||  
 Db 1957 Thr---PheGlnPro---ThrValAlaGlyTyrGlnThrSerTyrThrSerAsnAlaIle 1974  
 QY 1387 AATTTCTCTTCACTCGCTCTATGAGCGAGCCCATGAAGTCCCATATACAGGCCATC 1446  
 |||||  
 Db 1975 AsnIleAspValLeuThrValLysGlySerGlyHisPheValProLeuAspArgProGln 1994  
 QY 1447 GCTTCCCTGCAATGTTTAAACCGGACTATCTTGGTGGATATCGCAGAGGCCAGAG 1506  
 |||||  
 Db 1995 GlnAlaLeuGlnMet-----IleTyrAsnPheValLysSerArgGlyTyrAsn 2010  
 QY 1507 AAGATCTGGCCCGCTACAGACGAAATGAGCGGTACAGCTACCATACAGCTCGTCC 1566  
 |||||  
 Db 2011 ThrProPheAspLeuAsnSerAsnPheThrThrThrThrThrThrThrThrThrThr 2030  
 QY 1567 GTCCG-----CTGCCCTACGGCTACCGC 1590  
 |||||  
 Db 2031 ThrProGlyThrGlyProThrValThrAla 2040

## RESULT 14

Q9FMX9 PRELIMINARY; PRT; 499 AA.  
 ID Q9FMX9  
 AC Q9FMX9;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)



```

ID Q9FYP7 PRELIMINARY; PRT; 490 AA.
AC Q9FYP7;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Similar to Hordeum vulgare carboxypeptidase D (Putative
DE carboxypeptidase D).
GN P0011G08.38.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone:P0433F09.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone:P0011G08.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF002539; BAB08188.1; -.
DR HSSP; P08B19; 1WHT.
DR MPROPS; S10.005; -.
DR Gramene; Q9FYP7; -.
DR InterPro; IPR001563; Serine_carbpept.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR ProDom; PD001189; Serine_carbpept; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
DR Carboxypeptidase.
KW Carboxypeptidase.
SQ
SEQUENCE 490 AA; 54733 MW; 968FD450523D3574 CRC64;

Alignment Scores:
Pred. No.: 3,36e-19 Length: 490
Score: 385.00 Matches: 125
Percent Similarity: 41.77% Conservative: 68
Best Local Similarity: 27.06% Mismatches: 179
Query Match: 13.26% Indels: 90
DB: 10 Gaps: 19

US-09-712-338-1_COPY_55_1662 (1-1608) x Q9FYP7 (1-490)
QY 154 TACTCTGGATAT-----GTCGACACCTCCCGAGTCCCATACCTCTTCTGGTCTTC 207
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 56 TyrSerGlyTyrValThrValAspAlaAlaAlaGlyArgAlaLeuPheTyrTrpLeuIle 75
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 208 GAAGCAGACATACCCCAAAACGACCATTCACATTCGTTGGTGAATGGTGGCCCTGGA 267
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 76 GluAlaAla---AspProAlaSerAlaProLeuValLeuTrpLeuAsnGlyGlyProgly 94
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 268 AGCGATTCTTGTG---ATCGGCTCTTCGAGAGTTGGGCCCTTCCCATGTCATTCGACT 324
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 95 CysSerSerValGlyTyrGlyAlaSerGluGluLeuGlyAlaPheArgIleAsnProasp 114
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 325 TTTCATGAC---TACATCAACCCCTACTCGGGAACGAGGCTCCCAATTTACTATTCTG 381
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 115 GlyArgSerLeuTyrLeuAsnProTyrProTrpAsnArgValAlaAsnMetLeuPheLeu 134
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 382 TCCAGCCATGGGATGGCTTTTCATATAGTATGATGATGGTGGTCCATTAACCCCT 441
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 135 AspSerProAlaGlyValGlyTyrSerTyrSerAsnThrThrSerAspLeuPheThr--- 153
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 442 GTAACTGGGTCGTGGAATAATTCAGGCTTTCAGGAGTTCAGGGCCGTCACCAACCAT 501

```

```

Db 154 -----AlaGly----- 155
QY 502 GATGCCACCTGTATGATACCTACCAATCTTTCGCGAGAGCCGCTTGGAGATCCGCAA 561
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 156 -----AspAsnLysThrAlaHisAspSerTyrAlaPheLeuVal 168
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 562 GGATTCCTTAGTGACTACCTAGCTTGGAGCTCTPAGGCTCGAGCTAAGACTTCAGTCTA 621
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 169 AsnTrpLeuGluArgPhePro-----GluTyrLysTyrArgAspPheTyrIle 184
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 622 TGGACGAGAGCTATGAGGCGACTATGTCCTCGCATTCCTCAATCATTTTACGAGCAG 681
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 185 AlaGlyGluSerTyrGlyGlyHisTyrValProGlnLeuSerGlnLeuValTyrArgAsn 204
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 682 AATGAGAGAATTGCCAACGGTAGTGTATATGGTGTTCAGCTTAATTTCAACTCTCTGGGA 741
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 205 AsnLys-----AspValGluLysProIleLeuAsnPheLysGlyPheMet 219
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 742 ATTATTAAACGGCATCATCGACGAGCGATCCAGGCCCTTACACCTGATTCGCTGTG 801
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 220 ValGlyAsnAlaValIleAspAsp-----TyrHisAspTyrValGlyThrPheGluTyr 237
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 802 AACAAATACCTACGATCAAGGCTGTCAACGAGACCGCTACACACTACATCAAGTTTGC 861
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 238 TrpTrpThrHisGlyLeu-----IleSerAspThrTyrGlnLysLeuGlnValAla 255
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 862 AACCAAAATGCCAAAT-----GGTGGCCAGGAGTTG-----ATTTC 897
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 256 CysAspPheGluSerSerAlaHisAlaSerGluAlaCysAsnLysIleTyrGluValAla 275
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 898 ACCTGCAACAGACAAACCGCATTAAGTACTAGTACGCGCTCTCGCGGAGAGCAAC 957
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 276 GluAlaGluGlnGlyAsnIleAspAlaTyrSerIleTyrThrProThrCysLysThr 295
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 958 AACATGTGACGGACATGTTGAGGGCCATCTACGCCCTTGTGGTGGTGGTGTAT 1017
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 296 SerPheLeuLysArgArgLeuIleArgGlyAsnSerProTrpLeuProArgGlyTyr--- 314
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1018 GATATTGGCATCCATATGATGACCCGACTCCGCAAGTTATTATTAACAATTTCTGGCA 1077
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 315 -----AspProCysThrGluLysTyrSerThrLysTyrTyrAsn 327
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1078 AAGGACTCTGTATGAGCGCTATCGCGTCAACATCAAC-----TACACC 1122
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 328 LeuProGluValGlnLysAlaPheHisAlaAsnValThrGlyIleProTyrAlaTrpThr 347
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1123 CAGTCCATATAGCTACTAGCTTTCAGCAAAACAGCGACTTGTCTGCGCCCAAC 1182
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 348 ThrCysSerAspLeuPheTyrTyrTrpLysAspSerProArgSerMetLeuProIle 367
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1183 TTCATCGAAGACCTCGAGGAGATCCTTTCCTCCCGTGGTGTCTCCCTCATCTATGGC 1242
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 368 Tyr-----ArgGluLeuIleAlaGlyLeuArgIleTrpValPheSerGly 383
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1243 GACGCCGATTCATCTGCACTGGTTCGGCGCTCAGCCGCTTCCCTC---GCTCCGAAAC 1299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 384 AspAlaAspSerVal-----ValProLeuThrAlaThrArg 395
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1300 TACTCCCAAGCGCCCACTTCCGAGCGAGGTACAGCCCTCAAGAGTCAAGCGGCTC 1359
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 396 TyrSer-----IleAspAlaLeuTyrLeuProThrValThrAsnTrpTyr 410
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1360 GAGTAGGGGAAACTCGCGAG-----TATGGTAATTTCTCCTTC 1398
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 411 ProTrpTyrAspAspGluGluValAlaGlyTrpCysGlnValTyrGlnGlyLeuThrLeu 430
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1399 ACTCGCTCTATAGGCGAGCGCTCACTCCACTACTACCGCCCATCGCTCCCTCGCAA 1458
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 431 ValThrIleArgGlyAlaGlyHisGluValProLeuHisArgProArgGlnAlaLeuLys 450
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1459 TTGTTT 1464
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



Db 451 LenPhe 452

Search completed: September 16, 2003, 18:46:58  
Job time : 138.5 secs

**This Page Blank (uspto)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 16, 2003, 18:47:04 : Search time 61 seconds  
(without alignments)  
8368.273 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662

Perfect score: 536

Sequence: 1 cttccagggaagtacaccgcg.....gcattgtccagtggttgatg 1608

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Word size: 100

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=slp  
-Q=/cgn2\_1/USPTO.spool\_p/US09712338/runat\_16092003\_144403\_14541/app\_query.fasta\_1.1799  
-DB-A.Geneseq\_19Jun03 -OFMT=fastan -SUFFIX=n2poli.rag -MINMATCH=0.1 -LOOPCL=0  
-LIST=45 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=100 -ALIGN=15 -MODE=LOCAL  
-USER=US09712338 @CGN\_1\_1\_92.runat\_16092003\_144403\_14541 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

A.Geneseq\_19Jun03:\*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	535	99.8	554	19 AAW56099	A. oryzae ATCC2038

# ALIGNMENTS

## RESULT 1

AAW56099  
ID AAW56099 standard; Protein; 554 AA.

XX AAW56099;

XX 27-AUG-1998 (first entry)

XX A. oryzae ATCC20386 carboxypeptidase I protein.

XX Carboxypeptidase I; flavour improving agent; hydrolysate; proteinaceous; food industry.

XX Aspergillus oryzae.

XX Key Location/Qualifiers

FT Peptide 1..18

FT /label= signal

FT Protein 19..554

FT /label= carboxypeptidase I

XX WO9814599-A1.

XX 09-APR-1998.

XX 03-OCT-1997; 97WO-US17977.

XX 27-NOV-1996; 96US-0757534.

XX 04-OCT-1996; 96US-0726880.

XX (NOVO ) NOVO NORDISK BIOTECH INC.

XX (NOVO ) NOVO-NORDISK AS.

XX Berka R, Blinkovsky A, Brown K, Dambmann C, Golightly E;

XX Klotz A, Mathisen TE, Rey M;

XX WPI; 1998-240098/21.

XX N-PSDB; AAV28620.

XX Carboxypeptidase from Aspergillus oryzae - hydrolyses proteinaceous

XX substrates, useful for improving flavour of foods

XX Claim 1; Fig 3; 82pp; English.

XX This sequence represents carboxypeptidase I from Aspergillus oryzae.  
XX This polypeptide has an optimal activity in the range of pH 3.0-7.5 at  
XX 25 deg. C, optimal activity in the range of 55-60 deg. C at pH 4, and a  
XX residual activity of at least 65.5% after 30 minutes at pH 4.0 and  
XX 60 deg. C. It also has the capacity to hydrolyse X from N-CBZ-Ala-X where  
XX N-CBZ is N-carboxy and X is any amino acid. The carboxypeptidases  
XX can be used for obtaining hydrolysates (which can be enriched in free  
XX glutamic acid or peptide bound glutamic acid residues) from proteinaceous  
XX substrates. The carboxypeptidases can be used in flavour-improving  
XX compositions in the food industry. The products can also be used for the  
XX production of polypeptides free of carboxypeptidase activity.

XX Sequence 554 AA;

XX Alignment Scores:

Pred. No.: 0 Length: 554  
Score: 535.00 Matches: 536  
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00%  
 Query Match: 99.81%  
 DB: 19 Indels: 0  
 Gaps: 0

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x AAW56099 (1-554)

QY	1	CTTCCAGGAGTACACCGCGTCCGTCGGTAGAAGACAGTACCCAGAACCCACCGGG	60
Db	19	LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly	38
QY	61	GTCACAGCTCTACACCGGACAAATGTCACCATCCGGTACAGAACCCCGGGCAGAG	120
Db	39	ValLysThrLeuThrThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGlu	58
QY	121	GGGCTCTGGAGACTACCCCGGTGTCAATCTACTCTGGATATGTCGACACTCTCC	180
Db	59	GlyValCysGluThrThrProGlyValLysSerTyrSerGlyThrValAspThrSerPro	78
QY	181	GAGTCCCATACCTTCTCTGGTTCITTCGAAGCCAGACATAACCCAGAAATGCACTATC	240
Db	79	GluSerHisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIle	98
QY	241	ACATTCTGGTGAATGGTCCCTGGAAGGATTCTTGTATCGGTCTCTTCGAGAGTTG	300
Db	99	ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeu	118
QY	301	GGCCCTTGCCATGCAATTCGACTTTTGATGACTACATCAACCTCACTCGTGGACGAG	360
Db	119	GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGlu	138
QY	361	GTCCTCAATTTACTATTCCTGTCCAGCCATCGGAGTCGGCTTTTCATATAGTGATCG	420
Db	139	ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr	158
QY	421	GTTGATGGGTCCATTAACTGCTGAAGTGGGTGCTCCAAATTCGAGCTTTCGAGAGTT	480
Db	159	ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal	178
QY	481	CAGGGCCGGTACCAACCATGTGACCTGTCATGCTATGCTATGCTATGCTGCGCAGAG	540
Db	179	GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGlu	198
QY	541	GGCGCTGGGAGATCTGCAAGGATCTTAGTGGACTACCTAGCTTGGACTCTAGGGTG	600
Db	199	AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal	218
QY	601	CAGCTAAGGACTTCAGTCTATGGACGAGAGGATGAGGCGACATGCTGCTGCATTC	660
Db	219	GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhe	238
QY	661	TTCAATCATTTTACGACGAGAATGAGAGAATGCCAACGGTAGTGTATGTTGTTTCAG	720
Db	239	PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln	258
QY	721	CTTAATTTCACTCTCGGGAATTTAAGGGCAATCATCGACGAGGCGATCCAGGCCCT	780
Db	259	LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaPro	278
QY	781	TACTACCTGAATTCGCTGTGAACAATACCTACGGHATCAAGGCTGTCAACGACGACGTC	840
Db	279	TyrTrpProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrVal	298
QY	841	TACAACCTACATGAGTTTGGCCAAACAAATGCCAAATGGTTGCCAGGATTTGATTTCCACC	900
Db	299	TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr	318
QY	901	TGCAACAGACAAACCCGCGATTAGCTGACTACGCCCTCTCCGCGGAGGCCACCAAC	960
Db	319	CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn	338
QY	961	ATGTGACGGGACAATGTTAGGGGCCCATACTACGCCCTTGTGCTGGTGGTGGTGTATCAT	1020
Db	339	MetCysArgAspAsnValGluGlyProTyrThrAlaPheAlaGlyArgGlyValTyrAsp	358

Search completed: September 16, 2003, 18:58:12

Job time : 63 secs

QY	1021	ATTGCGCATCATATGATGACCCGACCTCCGCCAAGTTATTACAACTTCTGGCAAG	1080
Db	359	IleArgHisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys	378
QY	1081	GACTCTGTCTATGACGCTATCGGCTCAACATCACTACACCCAGTCCCAATATGAGTTC	1140
Db	379	AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal	398
QY	1141	TACTAGCTTTCCAGCAACAGCGGACTTTGTCTGGCCCACTTCATCGAAGACCTCGAG	1200
Db	399	TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu	418
QY	1201	GAGATCCTTCTCTCCCGCTCGGTGTCCTCATCTATGCGGACGCGGATCATCTGC	1260
Db	419	GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys	438
QY	1261	AACGTGTTCCGGCTCAGCGGCTTCTCCTCGCTCGAACHACTCCCAAGCGCCCACTTC	1320
Db	439	AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPhe	458
QY	1321	CGAAGCGCAGGGTACACGCGCCCTGAAAGTCAACGGCTCGAGTATGGGAAACTCGCGAG	1380
Db	459	ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu	478
QY	1381	TATGGTAATTTCTCCTTCACCTCGGCTCTATGAGGACGCCCATGAAGTCCCATACCAG	1440
Db	479	TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln	498
QY	1441	CCCATCGCCCTCCCTGCAATTGTTTAACCGGACTATCTTCGTTGGGATATCGCAGAGGC	1500
Db	499	ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGly	518
QY	1501	CAGAAGAAGATCTGGCCAGCTACAGACGAATGGAACGGCTACAGCTACGATACACAG	1560
Db	519	GlnLysLysIleIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGln	538
QY	1561	TGTPCCGTGGCGCTGCTGCTAGCGGTACAGCATGTCCTAGTGTGGTATG	1608
Db	539	SerSerValProLeuProThrAlaThrSerMetSerSerValGlyMet	554

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: September 16, 2003, 18:58:15 ; Search time 22 Seconds  
(without alignments)  
6185.078 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662  
Perfect score: 536  
Sequence: 1 cttccaggaggtacaccgac.....gcattccagtggttgatg 1608

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 328717 seqs, 42310858 residues

Word size: 100

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p\_model -DEV=xlp  
-Q=/cgn2\_1/USPro.spool\_P/US09712338/runat\_16092003.144405.14610/app\_query.fasta\_1.1/99  
-DB=Issued\_Patents\_AA -QFMT=fastan -SUFFIX=n2poli.rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=100 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09712338 -CGN 1 1 31 @runat\_16092003.144405.14610 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	535	99.8	554	3 US-08-943-714-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-943-714-2  
; Sequence 2, Application US/08943714  
; Patent No. 6187578  
; GENERAL INFORMATION:  
; APPLICANT: Blinkovsky, Alexander  
; APPLICANT: Berka, Randy

APPLICANT: Rey, Michael  
APPLICANT: Golightly, Elizabeth  
APPLICANT: Klotz, Alan  
APPLICANT: Mathisen, Thomas Erik  
APPLICANT: Dammann, Claus  
TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids  
TITLE OF INVENTION: Encoding Same  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6187578o No. 6187578disk of No. 6187578th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,714  
FILING DATE: 03-OCT-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4990.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 554 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-943-714-2  
Alignment Scores:  
Pred. No.: 0 Length: 554  
Score: 535.00 Matches: 536  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.81% Indels: 0  
DB: 3 Gaps: 0  
US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-08-943-714-2 (1-554)  
QY 1 CTTCCAGGAGGTACACCGCGCTCGGTAGACACAGCTACCCGAAGAACCCACCGGG 60  
Db 19 LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly 38  
QY 61 GTCAAGACCTTTACAACCGCAACAATGTCACCATCCGGTACAAAGAACCCCGGGCAG 120  
Db 39 ValLysThrLeuThrThrAlaAsnAsnValIleArgTyrLysGluProGlyAlaGlu 58  
QY 121 GGCGTCTGGGAGACTACCCCGGTGTCAATCTCTGATATGTCACACCTCTCCC 180  
Db 59 GlyValCysGluThrThrProGlyValLysSerTyrSerGlyThrValAspThrSerPro 78  
QY 181 GAGTCCCATACCTTCTTCTGGTTCCTTCGAGCCAGACATACCCAGAACTGCACCTAIC 240  
Db 79 GluSerHisThrPhePheTrpPhePheGluAlaArgHisAsnProGluThrAlaProIle 98  
QY 241 ACATTGTGGTTGAATGGTGGCCCTTGGAGAGCGATTCTTTGATCGGTCGAGAGTTG 300  
Db 99 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeu 118  
QY 301 GGCCCTTGCCATGTCAATTCGACTTTTGTATGACTACATAICAAACCCCTCACTCGTGAACGAG 360

```

119 GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGlu 138
120
361 GICTCCAAATTACTATTCCTGTCCTCCAGCCATTGGAGTCGGCTTTTCATATAGTATACG 420
139 ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 158
421 GTTGATGGTCCATTAACTTAACTGGGGTGGTCCGAAATTCGAGCTTTGAGGAGTT 480
159 ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal 178
481 CAGGGCCGTACCAACCACTTATGCTACTCTGATGATACATACCAATCTTCCCGCAGAG 540
179 GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGlu 198
541 GCCGCTTGGGAGATCCTCAAGATTCCTTACTGGACTACCTAGCTTGGACTCTAGGGTG 600
199 AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal 218
601 CAGTCTAAGGACTTCAGTCTATGGAGGAGAGCTATGGAGGCACTATGGTCTCGCATTC 660
219 GlnSerLysAspPheSerLeuThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 238
661 TTCATATTTTACGACGAGAATGAGAGAATTCGCAAGGTAGTGTAAATGGTGTTCAG 720
239 PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 258
721 CTTAATTTCACTCTCTGGGAATTAATTAACGGCATCATCGACGAGGCGATCCAGGCCCT 780
259 LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleAspGluAlaIleGlnAlaPro 278
781 TACTACCTGATTCGCTGTGCAACAATACCTACGGTATCAAGGTGTCAAGAGACCGGTC 840
279 TyrTyrProGluPheAlaValAsnThrTyrGlyIleLysAlaValAsnGluThrVal 298
841 TACAACACAGCAAAACCCGCGCATTAGCTGACTACGCCCTCTCGCGCGAAGCCACCAAC 960
319 CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn 338
961 ATGTGACGAGCACAAATGTTAGGGGCCATCTACCCCTTCTGCTGCTGGTGGTGTATCAT 1020
339 MetCysArgAspAsnValGluGlyProTyrIleAlaPheAlaGlyArgGlyValTyrAsp 358
1021 ATTCGGCATCCATATGATGACCGGCTCCGCGACTCCGCGAAGTTATWACAACAATTTCTGCAAG 1080
359 IleArgHisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys 378
1081 GACTCTGTCATGACCGCTATCGGGGTCAACATCACTACCCAGTCCCAATATAGCGTC 1140
379 AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal 398
1141 TACTAGCGCTTCCAGCAAAACAGCGACTTGTCTGCGCCCACTTCATCGAAGACCTCGAG 1200
399 TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGln 418
1201 GAGATCCTTGCTCTCCCGTGGTGTCTCCCTCATCTATGGCGAAGCCGATTACATCTGC 1260
419 GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys 438
1261 AACTGGTTCGGGGTTCAGCCGCTTCCCTGCTGCGAATCTACTCCCAAGCCGCCASITC 1320
439 AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPhe 458
1321 CGAAGCGCAGGTACACGCCCTGAAAGTCAACGGCTCGAGTATGGGGAACCTCCGAG 1380
459 ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyThrArgGlu 478
1381 TATGTATTTCTCCTCACTCGGCTCATGAGGAGGCGCATCAAGTCCCACTACACAG 1440
1440
```

Search completed: September 16, 2003, 19:04:29  
Job time : 24 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

No matches found

Search completed: September 16, 2003, 19:06:27  
Job time : 55 secs

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 16, 2003, 18:59:05 : Search time 55 Seconds  
(without alignments)  
8706.201 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662  
Perfect score: 536  
Sequence: 1 cttccaggaagtacaccgc.....gcattccagttgttgatg 1608

Scoring table: OLIGO

Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 556269 seqs, 148893369 residues

Word size: 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US09712338/runat\_16092003\_144405\_14644/app\_query.fasta\_1.1799  
-DB=Published Applications\_AA -QFMT=fastan -SUFFIX=n2poli.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=100  
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09712338.cgn\_1\_1\_15@runat\_16092003\_144405\_14644  
-NCPG=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications\_AA:

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score Match Length DB ID			Description
		Score	Match	Length	

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 16, 2003, 18:56:25 : Search time 36.5 Seconds

(without alignments)  
8473.383 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662

Perfect score: 536

Sequence: 1 ctccacgaagtagcacggc.....gcattgccagtgttggtatg 1608

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283308 seqs, 96168682 residues

Word size: 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPRO.spool\_P/US09712338/runat\_16092003\_144405\_14586/app\_query.fasta\_1.1799  
-DB=PIR\_76 -QMT=fastan -SUFFIX=n2poli.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=quality -THR\_MIN=100 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09712338\_@CGN\_1\_1\_57\_@runat\_16092003\_144405\_14586 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

No matches found

Search completed: September 16, 2003, 19:03:33  
Job time : 36.5 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 16, 2003, 18:48:44 : Search time 20.5 Seconds  
(without alignments)  
7377.458 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662  
Perfect score: 536  
Sequence: 1 cttccaggaggtacaccgc.....gcattccaggtgtgtgtatg 1608

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 127863 seqs, 47026705 residues

Word size: 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO\_scol\_p/US09712338/runat\_16092003\_144404\_14552/app\_query.fasta\_1.1799  
-DB=SwissProt\_41 -BPMI=fastan -SUFFIX=n2poli.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MAIRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=100 -ALIGN=15 -MODE=LOCAL  
-OUTEXT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09712338.cgn\_1.1.24.@runat\_16092003\_144404\_14552 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt\_41.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES			
Result	Query		
No.	Score	Match Length DB ID	Description
-----			

No matches found

Search completed: September 16, 2003, 18:59:01  
Job time : 20.5 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 16, 2003, 18:49:49 ; Search time 90 Seconds  
(without alignments)  
9221.080 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662  
Perfect score: 536  
Sequence: 1 cttcaggaagtacacggc.....gcattgctgattgtgtagt 1608

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Word size: 100

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2.1/USPTO\_pool\_P/US09712338/runat\_16092003\_144404\_14563/app\_query.fasta\_1.1799  
-DB=SPTREMBL\_23 -OFMT=fastan -SUFFIX=n2poli.rspt -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=100 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09712338 -CGN\_1\_123 -runat\_16092003\_144404\_14563 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_humani:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvrius:\*  
16: sp\_bacteriap:\*  
17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	434	81.0	555	3	Q96VZ9 aspergillus

#### ALIGNMENTS

##### RESULT 1

Q96VZ9  
ID Q96VZ9 PRELIMINARY: PRT; 555 AA.  
AC Q96VZ9;  
DT 01-DEC-2001 (TEMBLrel. 19, Created)  
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
DE Carboxypeptidase S1 precursor (EC 3.4.16.6).  
GN CPT.  
OS Aspergillus oryzae.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
RX NCBI\_TaxID=5062;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FK3;  
RA van den Broek P.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF394242; AAK77166.1; -  
DR InterPro: IPR001563; Serine\_carbpept.  
DR Pfam: PF00450; serine\_carbpept; 1.  
DR PRINTS; PR00724; CRBOXYPTASEC.  
DR ProDom; PD001189; Serine\_carbpept; 1.  
DR PROSITE; PS00131; CARBOXYPEPT\_SER\_SER; 1.  
KW Carboxypeptidase; Hydrolase; Signal.  
FT SIGNAL 1 18  
FT SIGNAL 18  
SQ SEQUENCE 555 AA: 61168 MW: 456563H0CEB55222 CRC64;

##### Alignment Scores:

Pred. No.: 0 Length: 555  
Score: 434.00 Matches: 535  
Percent Similarity: 99.81% Conservative: 0  
Best Local Similarity: 99.81% Mismatches: 1  
Query Match: 80.97% Indels: 0  
Gaps: 3  
US-09-712-338-1\_COPY\_55\_1662 (1-1608) x Q96VZ9 (1-555)

QY	1	CTTCAGGAAGTACACCGCGTCCGTCGTAAGACAGCTACCAAGAACCCACCGGG	60
Db	19	LeuProGlySerIhrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly	38
QY	61	GTCAGACTCTTACAAACGCAACATGTCACCATCCGTCACGAAGAACCCGGCAGAG	120
Db	39	ValLysThrLeuThrIhrAlaAsnValThrIleArgTyrLysGluProGlyAlaGlu	58
QY	121	GGCTCTCGGAGACTACCCGGGTGTCAAATCCTACTCTGGATATGTCGACACCTCTCC	180
Db	59	GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro	78
QY	181	GAGTCCCATACCTCTTCTGTTCTTGAAGCCAGACATAACCCAGAACTSCACCTATC	240
Db	79	GluSerHisThrPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIle	98
QY	241	ACATTGTGGTTGAATGTCGTCCTGGAAGCCATTCTTTGATCGGTCCTTCGAGAGTTG	300
Db	99	ThrLeuIrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeu	118
QY	301	GGCCCTGCCATGTCAATTCGACTTTTGATCATTACATCAACCCCTACTCGTGAACGAG	360
Db	119	GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGlu	138
QY	361	GTCTCCCAATTACTATTCTCTGCCAGCCATCGGACTCGGCTTTTCATATGATATACG	420
Db	139	ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerIyrAsnAspThr	158
QY	421	GTTCATGGTCCCATTAACCTGTAACTGGGGTCTCGAAATTCGAGCTTTGAGGAGTT	480
Db	159	ValAspGlySerIleAsnProValThrGlyValValGluAsnSerPheAlaGlyVal	178

Search completed: September 16, 2003, 19:02:12  
Job time : 92 secs

QY	481	CAGGGCGGTACCCACCAACATGATGCCACTCTGATCGATACCAATCTTGC	CGCGAGAG	540
Db	179	GlnGlyArgTy-ProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGlu	198	
QY	541	GCGCCTTGGAGATCCTCGAAGATTCCTTACTGGACTACCTAGCTTGGACTCTAGGGTG	600	
Db	199	AlaAlaIrrpGluLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal	218	
QY	601	CAGCTCAAGAGCTTCACTATGAGCGGAGAGCTATGAGGGCACTATGGTCCGCAATC	660	
Db	219	GlnSerLysAspPheSerLeuIrrpThrGluSerTyGlyGlyHisTyGlyProAlaPhe	238	
QY	661	TTCAATCATTTTACGAGCAGAAATGAGAGAAATGCCACGGTAGTGTGTTAATGGTCTTCAG	720	
Db	239	PheAsnHisPheTyGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln	258	
QY	721	CTTAATTTCAACTCTCTGGGAATTTAATACGGCATCATCGACGAGCGCATCCAGGCCCT	780	
Db	259	LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaPro	278	
QY	781	TACTACCTGAATTCGCTGTGAAACAATACCTACGGTATCAAGGCTGTCAACGAGACCGTC	840	
Db	279	TyrTy-ProGluPheAlaValAsnAsnThrTyGlyIleLysAlaValAsnGluThrVal	298	
QY	841	TACAACACTACATGAATTTGCCAACCAAAATGCCAAATGGTTCACAGATTTGATTTCCACC	900	
Db	299	TyrAsnTyrrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr	318	
QY	901	TGCAACACAGACAACCGCACCGCATTAGCTGACTACGCCCTCTGCGCGAAGCCACCAAC	960	
Db	319	CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrrAlaLeuCysAlaGluAlaThrAsn	338	
QY	961	ATGTGACGGACAAATGTTGAGGGGCCATACTACGCCCTTTGCTGGCTGGTGTGTATGAT	1020	
Db	339	MetCysArgAspAsnValGluGlyProTyrrAlaPheAlaGlyArgGlyValTyrrAsp	358	
QY	1021	ATTCGGCATCCATATGATGACCGCACCTCCGCCAAGTTATTACACAAATTTCTGGCAAG	1080	
Db	359	IleArgHisProTyrrAspAspProThrProProSerTyrrAsnLysPheLeuAlaLys	378	
QY	1081	GACTCTGTCATGGACGCTATCGGCGTCAACATCACTACACCCAGTCCAAATATGACGTC	1140	
Db	379	AspSerValMetAspAlaIleGlyValAsnIleAsnTyrrThrGlnSerAsnAspVal	398	
QY	1141	TACTACGCTTTCAGACAACAGCGACTTGTCTGCCCAACTTCATCGAAGACCTCGAG	1200	
Db	399	TyrTyrrAlaPheGlnGlnThrGlyAspPheValIrrProAsnPheIleGluAspLeuGln	418	
QY	1201	GAGATCCTTGCTCTCCCGTGGTCTCCCTCATCTATGGCGACCCGATACATCTCG	1260	
Db	419	GluIleLeuAlaLeuProValArgValSerLeuIleTyrrGlyAspAlaAspTyrrIleCys	438	
QY	1261	AACGTGTTGGCGGTACAGCCGTTTCCCTCGCTCGCACTACTCCCAAGCCGCCAGTTC	1320	
Db	439	AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrrSerGlnAlaAlaGlnPhe	458	
QY	1321	CGAAGCGAGGTACACGCCCTGAAAGTCAACGGCTCGAGTATGGGAAATCCGCGAG	1380	
Db	459	ArgSerAlaGlyTyrrThrProLeuLysValAsnGlyValGluTyrrGlyGluThrArgGlu	478	
QY	1381	TATGTAAATTTCTCTTCACTCGCGGTCTATGAGGCGAGCCCATGAHTCCCATCTACCAG	1440	
Db	479	TyrGlyAsnPheSerPheThrArgValTyrrGluAlaGlyHisGluValProTyrrTyrrGln	498	
QY	1441	CCCATCGCTCCCTCGCAATTTGTTAAACGGGACTATCTCGGTGGGATATCGCAGAGGC	1500	
Db	499	ProIleAlaSerLeuGlnPheAsnArgThrIlePheGlyIrrpAspIleAlaGluGly	518	
QY	1501	CAGAAGAAGATCTGGCCCGAGCTACAAGACGAATGGAACGGCTACAGCTACGATACACAG	1560	
Db	519	GlnIrrpValIleTrpProSerTyrrLysThrAsnGlyThrAlaThrAlaThrAsThrGln	538	

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 19:03:35 : Search time 7741 Seconds  
(without alignments)  
8497.956 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662  
Perfect score: 1608  
Sequence: 1 cttccaggaagtacaccgc.....gcattgtccagtgttggtatg 1608

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size : 100

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1608	100.0	1662	6	ARI29928	ARI29928 Sequence
2	531	33.0	2245	8	AF394242	AF394242 Aspergill

ALIGNMENTS

RESULT 1

ARI29928

LOCUS ARI29928 1662 bp DNA linear PAT 16-MAY-2001

DEFINITION Sequence 1 from patent US 6187578.

ACCESSION ARI29928

VERSION ARI29928.1 GI:14117825

KEYWORDS Unknown.

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1662)

AUTHORS Blinkovsky,A., Berka,R., Rey,M., Golightly,E., Klotz,A., Mathisen,T.Erik., Dambmann,C. and Brown,K.M.

TITLE Carboxypeptidases and nucleic acids encoding the same

JOURNAL Patent: US 6187578-A 1 13-FEB-2001;

FEATURES Location/Qualifiers

source l. 1662

BASE COUNT 396 a 468 c 398 g 400 t

ORIGIN

Query Match	100.0%	Score 1608;	DB 6;	Length 1662;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1608;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	CTTCAGGAGTACACCGGCTCGTGGTAGAGACAGCTACCCAGAACCCACCGGG	60	
Db	55	CTTCAGGAGTACACCGGCTCGTGGTAGAGACAGCTACCCAGAACCCACCGGG	114	
Qy	61	GTCAGACTCTTACAAACCGCAACAATGTCAACCTCGGTACAGAGAACCGGGGACAG	120	
Db	115	GTCAGACTCTTACAAACCGCAACAATGTCAACCTCGGTACAGAGAACCGGGGACAG	174	
Qy	121	GGGCTCTCGGAGACTACCCGGGTGTCAATCTTACTCTGGATATGTGGACACCTCTCCC	180	
Db	175	GGGCTCTCGGAGACTACCCGGGTGTCAATCTTACTCTGGATATGTGGACACCTCTCCC	234	
Qy	181	GAGTCCCAATACCTTCTTCTGTTCTTCCGAGCCAGACATACCCAGAACTGCACCTATC	240	
Db	235	GAGTCCCAATACCTTCTTCTGTTCTTCCGAGCCAGACATACCCAGAACTGCACCTATC	294	
Qy	241	ACATTGTGGTTGAATGGTGGCCCTGGAGCGAATCTTTGTATCGGTCTCTTCGAGAGTTG	300	
Db	295	ACATTGTGGTTGAATGGTGGCCCTGGAGCGAATCTTTGTATCGGTCTCTTCGAGAGTTG	354	
Qy	301	GGCCCTTGCCATGTCAATTTCGACTTTTGATGACTACATCAACCTCACTCGTGGAGAGG	360	
Db	355	GGCCCTTGCCATGTCAATTTCGACTTTTGATGACTACATCAACCTCACTCGTGGAGAGG	414	
Qy	361	GCTCCAAATTTACTATTTCCTGCTCCAGCCATTCGGAGTCGGCTTTTCATATAGTATACG	420	
Db	415	GCTCCAAATTTACTATTTCCTGCTCCAGCCATTCGGAGTCGGCTTTTCATATAGTATACG	474	
Qy	421	GTTGATGGGTCCCAATTAACCTGTAACCTGAGCTCGGAGTCGGCTTTTCATATAGTATACG	480	
Db	475	GTTGATGGGTCCCAATTAACCTGTAACCTGAGCTCGGAGTCGGCTTTTCATATAGTATACG	534	
Qy	481	CAGGCGCGGTACCCCAACCAATGTATGCCACTCTGTATCGATACCTACCAATCTTGCCGACAG	540	

Db 535 CAGGCGGCTACCAACCACTTATGATGACATCTATACCAATCTTGGCGCAG 594  
QY 541 GCGCTTGGAGATCCCTGCAAGATTCCTTAGTGGACTACCTAGCTTGGACTTAGGGTG 600  
Db 595 GCGCTTGGAGATCCCTGCAAGATTCCTTAGTGGACTACCTAGCTTGGACTTAGGGTG 654  
QY 601 CAGTCAAGGACTTTCAGTCTATGAGCGAGAGCTATGAGGCGCACTATGCTCCGCAATC 660  
Db 655 CAGTCAAGGACTTTCAGTCTATGAGCGAGAGCTATGAGGCGCACTATGCTCCGCAATC 714  
QY 661 TTCAATCATTTTACGAGCAGATGAGAGAAATGCCAACGGTAGTGTAAATGSGTGTTCAG 720  
Db 715 TTCAATCATTTTACGAGCAGATGAGAGAAATGCCAACGGTAGTGTAAATGSGTGTTCAG 774  
QY 721 CTTAATTTCACTCTCTGGGAATTTAAGCGCATCATCGAGGCGATCCAGGCCCT 780  
Db 775 CTTAATTTCACTCTCTGGGAATTTAAGCGCATCATCGAGGCGATCCAGGCCCT 834  
QY 781 TACTACCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACCGTC 840  
Db 835 TACTACCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACCGTC 894  
QY 841 TACAACATACATGAAGTTTGCAACCAAAATGATGATGATGATGATGATGATGATGATGAT 900  
Db 895 TACAACATACATGAAGTTTGCAACCAAAATGATGATGATGATGATGATGATGATGATGAT 954  
QY 901 TGCAACAGACAAACCGCAGCATACCTAGCTACGCCCTGCGCGGAGCCACCAAC 960  
Db 955 TGCAACAGACAAACCGCAGCATACCTAGCTACGCCCTGCGCGGAGCCACCAAC 1014  
QY 961 ATGTGCGAGGACAATTTGAGGGGCACTACCTACGCCCTTGTGCTGCTGCTGCTGCTGAT 1020  
Db 1015 ATGTGCGAGGACAATTTGAGGGGCACTACCTACGCCCTTGTGCTGCTGCTGCTGCTGAT 1074  
QY 1021 ATTCGGCATCATATGATGACCGGCTACCGGCTACCGGCTACCGGCTACCGGCTACCGGCT 1080  
Db 1075 ATTCGGCATCATATGATGACCGGCTACCGGCTACCGGCTACCGGCTACCGGCTACCGGCT 1134  
QY 1081 GACTCTGTCAGGAGCTATCGGCTCAACATCACTACCGGCTACCGGCTACCGGCTACCGGCT 1140  
Db 1135 GACTCTGTCAGGAGCTATCGGCTCAACATCACTACCGGCTACCGGCTACCGGCTACCGGCT 1194  
QY 1141 TACTACGCTTTCCAGCAACAGGCGACTTTGCTGCGCCCACTTTCATCGAAGACCTCGAG 1200  
Db 1195 TACTACGCTTTCCAGCAACAGGCGACTTTGCTGCGCCCACTTTCATCGAAGACCTCGAG 1254  
QY 1201 GAGATCCTTGTCTCCCGTGGTGTCTCCCTCATCTATGCGGAGCGCGATACATCTGC 1260  
Db 1255 GAGATCCTTGTCTCCCGTGGTGTCTCCCTCATCTATGCGGAGCGCGATACATCTGC 1314  
QY 1261 AACTGGTTCGGGGTCAGGCGGTTTCCCTGCTGGAATCTACTCCCAAGCCCGCATTC 1320  
Db 1315 AACTGGTTCGGGGTCAGGCGGTTTCCCTGCTGGAATCTACTCCCAAGCCCGCATTC 1374  
QY 1321 CGAAGCGAGGATACAGCGCCCTGAAAGTCAACGCGCTCGAGTATGGGAACCTCGGAG 1380  
Db 1375 CGAAGCGAGGATACAGCGCCCTGAAAGTCAACGCGCTCGAGTATGGGAACCTCGGAG 1434  
QY 1381 TATGTAATTTCTCTCTACTCGGCTATGAGGAGCGGCTATGAGGAGCGGCTATGAGGAGCGGCT 1440  
Db 1435 TATGTAATTTCTCTCTACTCGGCTATGAGGAGCGGCTATGAGGAGCGGCTATGAGGAGCGGCT 1494  
QY 1441 CCATCGCTCCCTCAATTTGTTTAAACGGACTATCTTCCGTTGGATATCGCAGAGGC 1500  
Db 1495 CCATCGCTCCCTCAATTTGTTTAAACGGACTATCTTCCGTTGGATATCGCAGAGGC 1554  
QY 1501 CAGAAGAAGATCGGCCAGCTACAGAGCAATGGAAGCGGTACAGCTACGATACACAG 1560  
Db 1555 CAGAAGAAGATCGGCCAGCTACAGAGCAATGGAAGCGGTACAGCTACGATACACAG 1614  
QY 1561 TCGTCGTCGCTGCTACGCGCTACGAGCATCTCCAGTGTGGTATG 1608

Db 1615 TCGTCGTCGCTGCGCTACGCTACGAGCAATGCCAGTGTGGTATG 1662  
RESULT 2  
AF394242 2245 bp DNA linear PLN 24-JUL-2001  
LOCUS Aspergillus oryzae strain TK3 carboxypeptidase SI (cpl) gene,  
DEFINITION complete cds.  
ACCESSION AF394242  
VERSION AF394242.1 GI:15004615  
KEYWORDS Aspergillus oryzae  
SOURCE Aspergillus oryzae  
ORGANISM Aspergillus oryzae  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
REFERENCE 1 (bases 1 to 2245)  
AUTHORS van den Broek, P.  
TITLE Direct Submission  
JOURNAL Submitted (22-JUN-2001) Bioscience, Nestec S.A., P.O. Box 44,  
Lausanne CH-1000, Switzerland  
FEATURES  
Location/Qualifiers  
1..2245  
/organism="Aspergillus oryzae"  
/mol\_type="genomic DNA"  
/strain="TK3"  
/db\_xref="taxon:5062"  
<1..>2245  
/gene="cpl"  
/product="serine carboxypeptidase 1"  
join(1..349,409..513,576..692,743..866,941..956,  
1008..1124,1187..1237,1296..1449,1522..1580,1670..>2245)  
/gene="cpl"  
join(1..349,409..513,576..692,743..866,941..956,  
1008..1124,1187..1237,1296..1449,1522..1580,1670..>2245)  
/gene="cpl"  
/EC\_number="3.4.16.6"  
/codon\_start=1  
/product="carboxypeptidase SI"  
/protein\_id="AAK77166.1"  
/db\_xref="GI:15004616"  
translation="MRGEFLSVLPVAASWALPGSTPASVGRRLPKNPTGKLTIT  
ANNVIRKEPGEAGVCTPGKSKYGVDPSPESHTEFFWFARHPETAPITLWL  
NGPGSDSLGLFEEELGCHVNSTEDDILNHSNVEYNLLFLOPLGVGSGYNPTVD  
SGINPTGVWNSFAGVGRYPTIDALITINLAARAAEILLOGLFSLDSRV  
QSKDFLSTESTYGHYAFENHYEQNERANGVQNLNGLSGLINGLIDRAIQ  
APYDFEAVNTYGIKAVNETVYMKFANQMGCDLITCKTNTATADYALCA  
EATNCRDNVEGYPVAFAGVYDIRHPDDPTPSYXKFLAKSVMDAIGVNTY  
QSNNDVYAFQGTGDFVNPFLDELEILALPVYSLLYGDADYICNKGGAVALAA  
NLSQAQRSAQYETELKVNVEYGETREYGNFETRVVEAGHEVYIYPIASLQLENR  
TIFGWDIAEGQKKIWFESYKNTGTATATHTQSSVPLPTATSSVSGWA"  
sig\_peptide 1..54  
BASE COUNT 548 a 603 c 524 g 570 t  
ORIGIN  
Query Match 33.0%; Score 531; DB 8; Length 2245;  
Best Local Similarity 100.0%; Pred. No. 9.2e-301;  
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1078 AAGGACTCTGTCATGACGCTATCGCGTCAACATCACTACACCCAGTCCCAATATGAC 1137  
Db 1709 AAGGACTCTGTCATGACGCTATCGCGTCAACATCACTACACCCAGTCCCAATATGAC 1768  
QY 1138 GTCTACTAGCTTTTCAGCAAAACAGGCGACTTTGTGTGCCCCAACTTCAICGAAGACCTC 1197  
Db 1769 GTCTACTAGCTTTTCAGCAAAACAGGCGACTTTGTGTGCCCCAACTTCAICGAAGACCTC 1828  
QY 1198 GAGGAGATCCTTGCTCTCCCGTGGTGTCTCCCTCATCTATGCGGACCGCATACATC 1257  
Db 1829 GAGGAGATCCTTGCTCTCCCGTGGTGTCTCCCTCATCTATGCGGACCGCATACATC 1888  
QY 1258 TGCAACTGCTTCGGCGCTCAGGCGGTTTCCCTGCTGCGAAGTCTACCTCCCAAGCGGCCAG 1317



Db 1889 TGCAACTGGTTCGGCGGTCTAGGCGGTTTCCTCGCTCGGAACACTCCCAAGCGGCCAG 1948  
QY 1318 TTCCGAAGCGCAGGGTACACGCCCTGAAGTCAACGGCTCGAGTAIGGGGAAACTCGC 1377  
Db 1949 TTCCGAAGCGCAGGGTACACGCCCTGAAGTCAACGGCTCGAGTAIGGGGAAACTCGC 2008  
QY 1378 GAGTATGGTAATTTCTCCTTCACTCGCGTCTATGAGCGAGCGCATGAAGTCCCATACTAC 1437  
Db 2009 GAGTATGGTAATTTCTCCTTCACTCGCGTCTATGAGCGAGCGCATGAAGTCCCATACTAC 2068  
QY 1438 CAGCCCATCGCCTCCCTGCAATGTTTAAACGGACTATCTCGGTGGGATATCGCAGAG 1497  
Db 2069 CAGCCCATCGCCTCCCTGCAATGTTTAAACGGACTATCTCGGTGGGATATCGCAGAG 2128  
QY 1498 GGCCAGAAGAAGATCTGGCCAGCTACAGACGANTCGGACGGCTACAGCTACGCATACA 1557  
Db 2129 GGCCAGAAGAAGATCTGGCCAGCTACAGACGANTCGGACGGCTACAGCTACGCATACA 2188  
QY 1558 CAGTCTCCGTGCGCGTCTACGGCTACCGGCTACCGCATGTCCAGTGTGGTATG 1608  
Db 2189 CAGTCTCCGTGCGCGTCTACGGCTACCGGCTACCGCATGTCCAGTGTGGTATG 2239

Search completed: September 16, 2003, 21:25:04  
Job time : 7741 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 19:06:30 ; Search time 148 Seconds  
(without alignments)  
4795.568 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662  
Perfect score: 1608  
Sequence: 1 cttccaggaagatcacacggc.....gcattgtccagtgttggtatg 1608

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 100

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCITUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1608	100.0	1662	3	US-08-943-714-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-943-714-1  
; Sequence 1, Application US/08943714  
; Patent No. 6187578  
; GENERAL INFORMATION:  
; APPLICANT: Blinkovsky, Alexander  
; APPLICANT: Berka, Randy  
; APPLICANT: Rey, Michael  
; APPLICANT: Golightly, Elizabeth  
; APPLICANT: Klotz, Alan  
; APPLICANT: Mathisen, Thomas Erik  
; APPLICANT: Dambmann, Claus  
; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids  
; TITLE OF INVENTION: Encoding Same  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6187578 No. 6187578disk of No. 6187578th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA

ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,714  
FILING DATE: 03-OCT-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4990.200-US  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1662 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1...1662  
OTHER INFORMATION:  
US-08-943-714-1

Query Match 100.0%; Score 1608; DB 3; Length 1662;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CTTCCAGGAAGTACACCGCGGTCGCGGTGAGGAAGACAGCTACCCCAAGAACCCACCGGG	60
DB	55	CTTCCAGGAAGTACACCGCGGTCGCGGTGAGGAAGACAGCTACCCCAAGAACCCACCGGG	114
QY	61	GTCAGAGCTTTACAAACCGCAACAAATGTACACATCCGGTACAGGAACCCGGGGCAGAG	120
DB	115	GTCAGAGCTTTACAAACCGCAACAAATGTACACATCCGGTACAGGAACCCGGGGCAGAG	174
QY	121	GGCGTCGCGAGACTACCCCGGGTGTCATCTCTGATATCTGACACCTCTCC	180
DB	175	GGCGTCGCGAGACTACCCCGGGTGTCATCTCTGATATCTGACACCTCTCC	234
QY	181	GAGTCCCATACCTTCTTCTGGTTCTTTCGAAGCCAGACATAACCCAGAAATGCACCTATC	240
DB	235	GAGTCCCATACCTTCTTCTGGTTCTTTCGAAGCCAGACATAACCCAGAAATGCACCTATC	294
QY	241	ACATTGTGGTGAATCGTGGCCCTGGAAGGATTCCTTGTATCGGTCTCTTCGAAGAGTTG	300
DB	295	ACATTGTGGTGAATCGTGGCCCTGGAAGGATTCCTTGTATCGGTCTCTTCGAAGAGTTG	354
QY	301	GGCCCTTGCCATGTCATTCGACTTTTGTATGACTATCAACCTCACCTCGTGGGAACGAG	360
DB	355	GGCCCTTGCCATGTCATTCGACTTTTGTATGACTATCAACCTCACCTCGTGGGAACGAG	414
QY	361	GTCTCCAAITTTACTATTCTTGTCCCGCATTCGGAGTGGGCTTTTCATATAGTATACG	420
DB	415	GTCTCCAAITTTACTATTCTTGTCCCGCATTCGGAGTGGGCTTTTCATATAGTATACG	474
QY	421	GTGTGAGGGTCCATTAACTGAACTGGGCTCGTGGAAATTCGAGCTTTTCAGAGTT	480
DB	475	GTGTGAGGGTCCATTAACTGAACTGGGCTCGTGGAAATTCGAGCTTTTCAGAGTT	534
QY	481	CAGGGCCGTACCAACCAATTTGATGCACTCTCATGATACCTATCTGCGGAGAG	540
DB	535	CAGGGCCGTACCAACCAATTTGATGCACTCTCATGATACCTATCTGCGGAGAG	594
QY	541	GCGGCTTGGGAGATCCCTTGTGAGATTCCTTAGTGGACTACCTAGCTTGGAGTG	600

Db 595 GCGGCTTGGAGATCTCGAAGATTCCTTAGTGGACTACCTAGCTGGACTAGGGTG 654  
QY 601 CAGCTAAAGACATTCAGTCTATATGAGCGGAGAGCTATGAGGGGCACTATGCTCTGCATTC 660  
Db 655 CAGCTAAAGACATTCAGTCTATATGAGCGGAGAGCTATGAGGGGCACTATGCTCTGCATTC 714  
QY 661 TTCATATCATTTTACGAGCAGAAATGAGAGAAATGCCAAGGTAAGTGTATATGGTGTTCAG 720  
Db 715 TTCATATCATTTTACGAGCAGAAATGAGAGAAATGCCAAGGTAAGTGTATATGGTGTTCAG 774  
QY 721 CTTAATTTCAACTCTCTGGGAATTTATTAAGGGCATCATCCAGAGCGGATCCAGGCCCT 780  
Db 775 CTTAATTTCAACTCTCTGGGAATTTATTAAGGGCATCATCCAGAGCGGATCCAGGCCCT 834  
QY 781 TACTACCCGAAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGAGCGGTC 840  
Db 835 TACTACCCGAAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGAGCGGTC 894  
QY 841 TACAACCTACATGAAGTTTGGCAACCAATGCCAAATGGTTGCCAGGATTTGATTTCCACC 900  
Db 895 TACAACCTACATGAAGTTTGGCAACCAATGCCAAATGGTTGCCAGGATTTGATTTCCACC 954  
QY 901 TGC AAAACAGACAAACCGCAGCATTAGCTGACTAGCCCTCTGCGCCGAGCCACCAAC 960  
Db 955 TGC AAAACAGACAAACCGCAGCATTAGCTGACTAGCCCTCTGCGCCGAGCCACCAAC 1014  
QY 961 ATGTGAGGAGCAATTTGAGGGGCCATATACGCGCTTTCGCTGGTGGTGTGTATGAT 1020  
Db 1015 ATGTGAGGAGCAATTTGAGGGGCCATATACGCGCTTTCGCTGGTGGTGTGTATGAT 1074  
QY 1021 ATTCGGCATCCATATGATGACCGGCTCCGCCAAGTTATTACAAACAAATTTCTGGCAAG 1080  
Db 1075 ATTCGGCATCCATATGATGACCGGCTCCGCCAAGTTATTACAAACAAATTTCTGGCAAG 1134  
QY 1081 GACTCTGTATGAGACCTATCGGCGTCAACATCACTACCCAGTCCCAATATATGACGTC 1140  
Db 1135 GACTCTGTATGAGACCTATCGGCGTCAACATCACTACCCAGTCCCAATATATGACGTC 1194  
QY 1141 TACTAGCGTTTCCAGCAACAGCGGACTTTGCTGCGCCCACTTCATCGAAGACCTCGAG 1200  
Db 1195 TACTAGCGTTTCCAGCAACAGCGGACTTTGCTGCGCCCACTTCATCGAAGACCTCGAG 1254  
QY 1201 GAGATCCTTGCTCTCCCGCTGGGTGTCCTCCCTCACTATGCGGACGCCGATACATCTGC 1260  
Db 1255 GAGATCCTTGCTCTCCCGCTGGGTGTCCTCCCTCACTATGCGGACGCCGATACATCTGC 1314  
QY 1261 AACTGGTTCGGGGTTCAGCGCTTTCCCTCGCTGCGAATCTCCCAAGCGCCCACTTC 1320  
Db 1315 AACTGGTTCGGGGTTCAGCGCTTTCCCTCGCTGCGAATCTCCCAAGCGCCCACTTC 1374  
QY 1321 CGAAGCGCAGGATACAGCCCTCGAAAGTCAAGCGGCTGAGTATGGGGAACCTCGCGAG 1380  
Db 1375 CGAAGCGCAGGATACAGCCCTCGAAAGTCAAGCGGCTGAGTATGGGGAACCTCGCGAG 1434  
QY 1381 TATGTAATTTCTCTCTACTCGGCTATGAGGAGGCCATGAGTCCCATACACAG 1440  
Db 1435 TATGTAATTTCTCTCTACTCGGCTATGAGGAGGCCATGAGTCCCATACACAG 1494  
QY 1441 CCCATCGCCTCCCTGCAATTTTAAACGGGACTATCTTCGGTTGGGATATCGCAGAGGC 1500  
Db 1495 CCCATCGCCTCCCTGCAATTTTAAACGGGACTATCTTCGGTTGGGATATCGCAGAGGC 1554  
QY 1501 CAGAAGAGATGCGCCAGCTACAGAGCAATGGAACGGCTACAGCTACGCATACACAG 1560  
Db 1555 CAGAAGAGATGCGCCAGCTACAGAGCAATGGAACGGCTACAGCTACGCATACACAG 1614  
QY 1561 TCGTCCGTCGCTGCTACGCTACCGGATCCAGCATCTCCAGTTGGTATG 1608  
Db 1615 TCGTCCGTCGCTGCTACGCTACCGGATCCAGCATCTCCAGTTGGTATG 1662

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 19:16:01 : Search time 517 Seconds  
(without alignments)  
7650.964 Million cell updates/sec

Title: US-09-712-338-1\_copy\_55\_1662  
Perfect score: 1608  
Sequence: 1 cttccaggaagtacaccgc.....gcatgtccagtgtgtatg 1608

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1660708 seqs, 1229959015 residues

Word size : 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications\_NA:\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/PCRS\_PUBCOMB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09C\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US50\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
-----					

No matches found

Search completed: September 16, 2003, 22:46:25  
Job time : 517 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

Search completed: September 16, 2003, 22:34:58  
Job time : 4184 secs

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 19:04:35 ; Search time 4184 Seconds  
(without alignments)  
9340.726 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662  
Perfect score: 1608  
Sequence: 1 cttccaggaagatcacccgc.....gcattgcagtggtgtatg 1608

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152236056 residues

Word size : 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST.\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pla:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gssl:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

No matches found

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: September 16, 2003, 14:42:53 ; Search time 7740 Seconds  
(without alignments)  
8499.054 Million cell updates/sec  
Title: US-09-712-338-1\_COPY\_55\_1662  
Perfect score: 1608  
Sequence: 1 cttccaggaaagtacaccgc.....gcattgcaggttggtatg 1608  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2888711 seqs, 20454813386 residues  
Total number of hits satisfying chosen parameters: 5777422  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1608	100.0	1662	6	ARI29928	ARI29928 Sequence
2	937.8	58.3	2245	8	AF394242	AF394242 Aspergill
3	874.8	54.4	1656	6	AX534871	AX534871 Sequence
4	313.6	19.5	3150	6	AX534814	AX534814 Sequence
5	134.4	8.4	1872	6	AX534872	AX534872 Sequence
6	118.6	7.4	1665	6	AX534866	AX534866 Sequence
7	116.8	7.3	1581	6	AX534824	AX534824 Sequence
8	116.8	7.3	2940	6	AX534767	AX534767 Sequence
9	114.6	7.1	3080	6	AX534809	AX534809 Sequence
10	107.8	6.7	3221	6	AX534815	AX534815 Sequence
11	86.4	5.4	2503	6	A75535	A75535 Sequence 2
12	86.4	5.4	2503	6	I70282	I70282 Sequence 3
13	84	5.2	2002	6	I33983	I33983 Sequence 3
14	84	5.2	2002	6	I74375	I74375 Sequence 3
15	84	5.2	2002	6	I77239	I77239 Sequence 3
16	82.4	5.1	1611	6	AX534826	AX534826 Sequence
17	78.2	4.9	2660	6	AX534769	AX534769 Sequence
18	76	4.7	2068	6	I33982	I33982 Sequence 1
19	76	4.7	2068	6	I74374	I74374 Sequence 1
20	76	4.7	2068	6	I77238	I77238 Sequence 1
21	74.6	4.6	3850	6	E12103	E12103 DNA encodin
22	74.6	4.6	3850	8	PPRC1GEN	X87987 P.pastoris
23	72.4	4.5	660	8	CNS01BTO	AL114676 Botrytis
24	70.6	4.4	2214	8	PA067174	U67174 Pichia angu
25	68.4	4.3	1368	6	AX534850	AX534850 Sequence
26	68.4	4.3	1653	6	AX536918	AX536918 Sequence
27	68.4	4.3	2441	6	AX534793	AX534793 Sequence
28	68.2	4.2	4308	8	D86560	D86560 Schizosacch
29	68.2	4.2	37000	8	SPAC19G12	Z97209 S.pombe chr
30	67.6	4.2	1985	8	YSACARPEPY	M95182 Candida alb
31	66.2	4.1	2632	8	ARI29959	ARI29959 Sequence
32	66.2	4.1	2632	8	YSCPRCCPY	M15482 Yeast vacuo
33	66.2	4.1	1581	8	SC8175	X80836 S.cerevisia
34	65.6	4.1	2280	8	AB051820	AB051820 Aspergill
35	65.6	4.1	2509	8	AF085063	AF085063 Pichia an
36	63.2	3.9	1446	6	AX534852	AX534852 Sequence
37	63.2	3.9	2520	6	AX534795	AX534795 Sequence
38	61.8	3.8	1527	6	AX594884	AX594884 Sequence
39	61.8	3.8	2016	8	SCYBR139W	Z36008 S.cerevisia
40	61.8	3.8	2027	6	AX536454	AX536454 Sequence
41	61.8	3.8	29686	8	SC29711	X75891 S.cerevisia
42	61.8	3.8	50277	2	ACL138524_5	Continuation (6 of
43	59.6	3.7	1580	8	HVACXPL12	X78878 Hordeum vul
44	59.4	3.7	4760	3	MOSAOR	L46594 Aedes aegypt
45	59.2	3.7	1664	3	AK115283	AK115283 Ciona int

ALIGNMENTS

RESULT 1	ARI29928	Sequence 1	1662 bp	DNA	linear	PAT 16-MAY-2001
LOCUS	ARI29928	Sequence 1 from patent US 6187578.				
DEFINITION	ARI29928					
ACCESSION	ARI29928.1	GI:14117825				
VERSION	ARI29928.1					
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 1662)					
AUTHORS	Blinkovsky,A., Berka,R., Rey,M., Golightly,E., Klotz,A., Mathisen,T.Erik., Dambmann,C. and Brown,K.M.					
TITLE	Carboxypeptidases and nucleic acids encoding the same					
JOURNAL	Patent: US 6187578-A 1 13-FEB-2001;					



```

/gene="cpi"
/product="serine carboxypeptidase 1"
join(1..349,409..513,576..692,743..866,941..956,
1008..1124,1187..1237,1296..1449,1522..1580,1670..2245)
/gene="cpi"
/EC_number="3.4.16.6"
/codon_start=1
/product="carboxypeptidase s1"
/protein_id="AAK7166.1"
/db_xref="GI:15004616"
/translation="MRGYEFLSVLPLVAASWALPGSTPASVGRRLKPNKPTGVKTLIT
ANNVTRYKPGAEVCGETIPGVKSYGYVDIIPESHFFWFEEARHNPTAPITLWL
NGGPGSLLGLEEFELCPCHVNSTFDYINPHSWNEYSNLLFLSOPLGVEFSYNDVD
GSINPTGVVSESGFACQGPYDIDATIDFTNLAAEAAWEIIQGLSLPLSDRSV
QSKDFSLWTEYSGHGVPAFENHFEONERIANGSVNGVNLNSLITINGLIDEALQ
APVYPEAVANTYGIKAVNETVYNYKKFANMENGCDLLSTCKQTKRNLALALCA
EATNMRDNVEGPIYAFAGRGVYDRIHPYDDPTPPSYNKLFLAKDSYMDAIGVINNT
QSNVDYAFQOQGFDFPNFIEDLEEILALPVRSIYGLADYICMWFQGVASLAA
NYSQAQFRSAGYTPLVKNGVEYGETREYGNFSFTRYVEAGHEVPYQPIASLQLENR
TIFGWDIAEGKKIWPYKNGKNGTATATHTQSSVPLPTATSMSSVGM"
sig_peptide 1..54
BASE COUNT 548 a 603 c 524 g 570 t
ORIGIN

Query Match      58.3%; Score 937.8; DB 8; Length 2245;
Best Local Similarity 73.5%; pred. No. 1e-231;
Matches 1606; Conservative 0; Mismatches 2; Indels 577; Gaps 9;

QY 1  CTTCCAGGAAGTACACCGCGCTCGGTGGTAGAGACAGTACCCAAAGAACCCACCGGG 60
Db 1  CTTCCAGGAAGTACACCGCGCTCGGTGGTAGAGACAGTACCCAAAGAACCCACCGGG 114
QY 61  GTCAAGACTCTTACAACCGCAACAATGTCAACATCGGTTACAGGAACCCGGGGAGAG 120
Db 115  GTCAAGACTCTTACAACCGCAACAATGTCAACATCGGTTACAGGAACCCGGGGAGAG 174
QY 121  GCGGTCTGGGAGACTACCGCGGTGTCAAACTCTACTCTGGATATGTCGACACTCTCC 180
Db 175  GCGGTCTGGGAGACTACCGCGGTGTCAAACTCTACTCTGGATATGTCGACACTCTCC 234
QY 181  GAGTCCCATACCTCTCTCTGTTCTTTCGAAGCCAGACATAACCCAGAACTGCACCTATC 240
Db 235  GAGTCCCATACCTCTCTCTGTTCTTTCGAAGCCAGACATAACCCAGAACTGCACCTATC 294
QY 241  ACATTGTGTTGAATGTGTCCTCGGACGATCTTGTGATCGTCTCTCGA----- 293
Db 295  ACATTGTGTTGAATGTGTCCTCGGACGATCTTGTGATCGTCTCTCGAAGGTTG 354
QY 294  -----AGAGTTGG 301
Db 355  CTTGACGCTGTTACAAGTATGCTCTTTTAGGTCGGTTAACCTTTGTTAGAGTTGG 414
QY 302  GCGCTTGCATGCAATTCGACTTTGATGATACATACCTCTACTCTGATATGTCGACACTCTCC 361
Db 415  GCGCTTGCATGCAATTCGACTTTGATGATACATACCTCTACTCTGATATGTCGACACTCTCC 474
QY 362  TCTCCAAATTTACTATTCCTCTCCAGCCATTTGGAGTC----- 399
Db 475  TCTCCAAATTTACTATTCCTCTCCAGCCATTTGGAGTCGGTATGGAATGCGCGACCTCC 534
QY 400  -----GGCTTTTCATATAGTATAC 419
Db 535  CCTACTGGGTCTTCCCAATTTGACGAGGTGCTTCCCGTAGGCTTTTCATATATGATAC 594
QY 420  GGTGTATGGTGCATTAACCTCTGAATGAGTCTGTCGAAATTCGAGCTTTTCGAGGAGT 479
Db 595  GGTGTATGGTGCATTAACCTCTGAATGAGTCTGTCGAAATTCGAGCTTTTCGAGGAGT 654
QY 480  TCAGGCCGCGTACCCAAACATTTGATGCCACTCTGATC----- 516
Db 655  TCAGGCCGCGTACCCAAACATTTGATGCCACTCTGATCGGTAAAGGTTTCTCTCGGTCCAG 714

```

```

QY 517  -----CATACIACCAATCTTTGCCGACAGAGGCCCTTGG 549
Db 715  ATGCTTCTATTACTGATACGATAGATACCAATCTTCCGACAGAGGCCCTTGG 774
QY 550  GAGATCCTGCAAGGATTCCTTAGTGGACTACTAGCTTGGACTCTAGGCTCAGTCTAAG 609
Db 775  GAGATCCTGCAAGGATTCCTTAGTGGACTACTAGCTTGGACTCTAGGCTCAGTCTAAG 834
QY 610  GACTCAGTCTATGACGAGGAGCTATGGAG----- 640
Db 835  GACTCAGTCTATGACGAGGAGCTATGAGGAGTGGTTCCTCTCTCGAATCTTCA 894
QY 641  -----GGCACTATGGTCTCTG 655
Db 895  AACGGCCGTGATGTTGGGTGATAGGTTTCTTTTCGTTGAAAGGACATATGGTCTG 954
QY 656  CA-----TTCATCA 664
Db 955  CAGTTCAGTCTACCTGGACGATGTGTCAAAAGCAAGCACTAAGCAAGTAGTCTTCA 1014
QY 665  ATCAATTTTACGACGAGAAATGTCACACGCTAGTGTAAATGGTGTTCAGCTTA 724
Db 1015  ATCAATTTTACGACGAGAAATGTCACACGCTAGTGTAAATGGTGTTCAGCTTA 1074
QY 725  ATTTCAACTCTCTGGAAATTTAAACGGCATCATCGAGGCGATCC----- 772
Db 1075  ATTTCAACTCTCTGGAAATTTAAACGGCATCATCGAGGCGATCCAGGTATACCGTC 1134
QY 773  -----AGGCCCTTCA 782
Db 1135  GAAACAGCCCATAGAACATCTCCGACAAAGCTGTAACTATGCCATAGGCCCTTCA 1194
QY 783  CTACCTCGAATTCGCTGTGAACAATACCTACGGTATCAAGGCT----- 825
Db 1195  CTACCTCGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTTAAGTTCCCTTATG 1254
QY 826  -----GTCAAGCAGACCGTCTACA 844
Db 1255  CGCATGAGATTATATCTCTGATACTCATCACCATGAAAGGTCAACGAGACCGTCTACA 1314
QY 845  ACTACATGAAGTTTGCACCAACAAATGTCACAGGATTTGATTTCCACCTGCA 904
Db 1315  ACTACATGAAGTTTGCACCAACAAATGTCACAGGATTTGATTTCCACCTGCA 1374
QY 905  AACAGAAACCGCACCGCATTTAGCTACAGCCCTCTCGCCGAGAACCAACCAATGT 964
Db 1375  AACAGAAACCGCACCGCATTTAGCTACAGCCCTCTCGCCGAGAACCAACCAATGT 1434
QY 965  GCAGGGACAATGTT----- 978
Db 1435  GCAGGGACAATGTTGGTACGTTCTTTTCCCTCTTGGAAACACAGGTATATGATCAAAA 1494
QY 979  -----GAGGGGCCATACAGCCCTTTGGCTTGGTGGTGGTGG 1012
Db 1495  TAGTGGGACAATAACACGAAAAATAGAGGGCCATACAGCCCTTTGGCTTGGTGGTGG 1554
QY 1013  TGTATGATTTTGGCATCCATATGAT----- 1038
Db 1555  TGTATGATTTTGGCATCCATATGATGTAAGTAGAGCTTTTACTTCTATTATTACACCGC 1614
QY 1039  -----GACCC 1043
Db 1615  ATGCAATGATGGTTCATGTTGATCGAATAAAATCCOCCCTAATATGTTATAGGACCC 1674
QY 1044  GACTCCGCCAAGTTATTACAACAAATTTCTGGCAAGGACTCTGTGATGACCTATCGG 1103
Db 1675  GACTCCGCCAAGTTATTACAACAAATTTCTGGCAAGGACTCTGTGATGACCTATCGG 1734
QY 1104  CGTCAACATCAACTACACCCAGTCCCAATATGACGCTCTACTAGCTTTCCAGCAACAGG 1163
Db 1735  CGTCAACATCAACTACACCCAGTCCCAATATGACGCTCTACTAGCTTTCCAGCAACAGG 1794
QY 1164  CGACTTTGTCTGCCCAACTTCATCGAAGACCTTCGAGGAGATCCCTTCTCTCCCGTGG 1223

```

```

1795 CGACTTTGCTGGCCCACTTCATCGAAGACCFCGAGGAGATCCTTTGCTCTCCCGCTGGC 1854
1224 TGTCTCCCACTATATGCGGACGCCGATATACATCTGCAACTGCTTGGCGGCTCAGGCGGT 1283
1855 TGTCTCCCTCATCTATGCGGACGCCGATATACATCTGCAACTGCTTGGCGGCTCAGGCGGT 1914
1284 TTCCCTCGCTGGCACTACTCCCAAGCGCCCAAGTTCGGAAGCGCAGGFTACAGCGCCCT 1343
1915 TTCCCTCGCTGGCACTACTCCCAAGCGCCCAAGTTCGGAAGCGCAGGFTACAGCGCCCT 1374
1344 GAAAGTCAACGGCGCTGAGTATGCGGAAACTCGCGAGTATGTTAAATTTCTCCCTCACTCG 1403
1975 GAAAGTCAACGGCGCTGAGTATGCGGAAACTCGCGAGTATGTTAAATTTCTCCCTCACTCG 2034
1404 CGTCTATGAGCGAGCCATCAAGTCCCATCTACACGCCCATCGCCCTCCCTGCAATGTT 1463
2035 CGTCTATGAGCGAGCCATCAAGTCCCATCTACACGCCCATCGCCCTCCCTGCAATGTT 2094
1464 TAACCGGACTATCTCGGTGGGATATCGAGAGGGCCGAGAAAGATCTGGCCCGAGCTA 1523
2095 TAACCGGACTATCTCGGTGGGATATCGAGAGGGCCGAGAAAGATCTGGCCCGAGCTA 2154
1524 CAAGACGAATGAGCGCTACAGCTACGCATACACATACAGTCTCGTGGCGCTGCTCAGGC 1583
2155 CAAGACGAATGAGCGCTACAGCTACGCATACACATACAGTCTCGTGGCGCTGCTCAGGC 2214
1584 TACCAGCATCTCCAGTGTGGTATG 1608
2215 TACCAGCATCTCCAGTGTGGTATG 2239

```

## RESULT 3.

```

AX534871 LOCUS AX534871 1656 bp DNA linear PAT 22-NOV-2002
DEFINITION Sequence 109 from Patent WO02068623.
ACCESSION AX534871
VERSION AX534871.1 GI:25261462
KEYWORDS
SOURCE Aspergillus niger
ORGANISM Aspergillus niger

```

```

REFERENCE 1
AUTHORS Edens,L., dijk Van,A.A., Krubasik,P., Albermann,K., Stock,A.,
Kimpel,E., Klugbauer,S., Wagner,C., Fritz,A., gustedt Von,W.,
Heinrich,O., Maier,D., Spreafico,F., Folkers,U., Hopper,S.,
Kemmer,W., Tan,P., Stiebler,J. and Albang,R.
Novel genes encoding novel proteolytic enzymes
Patent: WO 02068623-A 109 06-SEP-2002;
DSM N.V. (NL)

```

```

FEATURES             Location/Qualifiers
     source            1..1656
                        /organism="Aspergillus niger"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:5061"
BASE COUNT            411 a 426 c 409 g 410 t
ORIGIN

```

```

Query Match          54.4%; Score 874.8; DB 6; Length 1656;
Best Local Similarity 71.7%; Pred. No. 2.1e-215;
Matches 1146; Conservative 0; Mismatches 452; Indels 0; Gaps 0;

```

```

QY      2 TTCCGAGGAGTACACCGCGCTCGGTGAGACAGACAGCTACCCAAAGAACCCACCGGG 61
DB      56 TCCCGAGATGATGTGTCATCTACGATAGAGAGCGAGTTACCAAAACGCTCCACTGGC 115
QY      62 TCAAGACTCTTACAAACCGCAACAAATGTACCAATCCGGTACAAAGAAACCGGGGACAGG 121
DB      116 TCAATCGATAAAACCCCAACAAATGTCACTATCAGTATATAAAGAACCAAGAACCGAAG 175
QY      122 GCGTCTGGGACACTACCCCGGGTGTCAAACTCTACTCTGATATGTGACACCTCTCCG 181

```

```

176 GAAATTTGAGACAACACCTGGGGTCAAAATCAATACCTCCGGAATATGTGATCTTTCCGCAG 235
182 AGTCCCATACCTTCTTCTGTTCTTGAAGCCAGACATAACCCAGAAACTGCACCTATCA 241
236 AGTCGATACATTTCTTTTGGTTTTCGAGTACGCGTGAACCGGAAATGATCCAGTGA 295
242 CATTTGGTTGAATGTTGGCCCTGGGAAGCGATCTTTGATCGGTTCTTCCAGAGTTGG 301
296 CTCGTGGCTGAATGTTGGCCCTGGGAAGCGATCTTTGATCGGTTCTTTTGAAGAGTTGG 355
302 GCCCTTGGCAATGATCAATTCACATTTTGATGACTACATCAACCTCACTCGTGGAAACAGG 361
356 GTCGGTGTACATCACACACAGAGTACGAATCAATCAATCACTACTCTCTGGAACAGG 415
362 TCTCCAAATTTACTATTTCTCTCCAGCCATTTGGGAGTCGGCTTTTTCATATAGTATACGG 421
416 TCACCAATCTTCTTCTGTCACGCCCTCGGTGTTGGGTCTCTTACAGTGAACCG 475
422 TTGATGGGTCCATTAACCCCTGTAACCTGGGTGTCGGAATTCGAGCTTTGCAAGAGTTC 481
476 AGCCGGGTCTTGAATCCATTTACTGAGCGCTGAGAGCGCTCCCTTTGCTGGAGTTC 535
482 AGGCCGGTACCCAAACCATTCATGSCCACTCTGATCGATACTACCAATCTTGCCGCAGAG 541
536 AGGTGATACCCAGTTATTGATGCCACTATCATCGACACGCCGATATGCTGCACCG 595
542 CCGCTTGGGAGATCTGCAAGGATTCCTTAGTGACTACCTAGCTTGGACTCTAGGTCG 601
596 CAACCTGGAGGTCTTCAGGGCTTCCTCAGTGGCTTCCAGCTGTCGACGTAGATTCGGAAGTCA 655
602 AGTCTAAGGACTTCAGTCTATGAGCGGAGACTATGAGGCGCAGTATGGTCTGCTCATCT 661
656 AGTCCAAAGAGTTCAACCTGTGGACAGAGTATAGGAGGACACTATGGACAGCGTCT 715
662 TCANTCATTTTACGACAGCAATGAGAAATGTCACAGGTAGTTAAATGCTGTTCAGC 721
716 TCAATCATTTCTACGACGAAATTCGAAGATCGCTAGCGGGAGTAGTCATGCGCTCCAC 775
722 TTAATTTCACTCTCTGGGAATTAATAACGGCATATCGACGAGCGCATCCAGGCCCTT 781
776 TGAATTTAATCTCCCTCGGATTAACGSCATCATTTGATGCGCGCATTCAGCAGACT 835
782 ACTACCTGAAATTCGCTGTGAACAATACCTACGATATCAAGGCTGTCAACAGACCGTCT 841
836 ACTACGAGACTTTTCCGCTTAATAATACATATGAATCAAAASCTGTCAATACACACTGT 895
842 ACAACTACATCAACTTTGCCAACCAATGCCAAATGTTGCCAGGATTTGATTTCCACT 901
896 ACAACTATATGAGTTCCGCCAACACGATGCCAAATGATGCCAGATCAGGTTGCTGCT 955
902 GCAAAACAGACAAACCGCACCGCATTAGCTAGCTAGCGCTCTGCGCGAAGACCCCAACA 961
956 GTAAATTGACCAATAGGACCTCGCTTCTGATATATGATATATGATACAGACGACCAATA 1015
962 TGTGAGGACAATTTGAGGGGCCATACACTAGCCCTTTGCTGGTGGTGTGATGATA 1021
1016 TGTGAGGACAATTTGAGGGGCCCTTACTACAGTTTGGCGGGCGTGGCGTGTATGATA 1075
1022 TTGSGATCCATATGATGACCCGACTCCGCAAGTTATTAACAACAAATTTCTGCGCAAGG 1081
1076 TTGCGACCCCTCAATGACCCGACCCCGCGTCTACTTTGTTCACTACCTCAAGAAAG 1135
1082 ACTCTGTATGAGCGCTATCGCGGTCAACATCAACTACACCGAGTCCCAATATGACGCT 1141
1136 ACTCAGTCATGGATGCTATCGCGGTGGACATTAACATACCGGAGTCCAGCGCGAAATAT 1195
1142 ACTAGCTTTCCAGCAACACAGCGACTTTGTCGCCCAACTTCATCGAAGACCTTCGAGG 1201
1196 AATATGCAATTCAGACAGACCGCGACTTTGATGGCCGAATTTCAATGAGACCTTCGAG 1255
1202 AGATCTGTCTCCCGGTGGTGTCTCCCTCATCTATGGGACCGCCGATTAATCATCTGCA 1261
1256 AGATCTCCAACTCCCGGTAGCGGTGCTGATCTACGGGATCGGACTATATCTGTA 1315

```

Qy	1262	ACTGGTTGCGGGTCAGSCGGTTTCCCTCGCIGCGAACTACTCTCCAAAGCGCCGAGTTC	1321
Db	1316	ACTGGTTGCGGGTCAGSCCATCTCACTCGAGTTACTACCCCATCGAGCTCAGTTC	1375
Qy	1322	GAAGCGGAGGGTACACGCCCTCTCAACGGCGTCGAGTAGTGGGAAACTCGCGAGT	1381
Db	1376	GTGACGCGGATACACACCATACAGATAGATGGGTGGAATACGGTGAAGTTCGCGAGT	1435
Qy	1382	ATGTAATTTCCTTTCACTCGGGTCTATGAGCGAGGCCATGAAGTCCCATACTACGAGC	1441
Db	1436	ATGCAACTTTTCGTTACCCGGGTATATCAGGTGGGCAGAGTTCCATATCAAC	1495
Qy	1442	CCATCGCCTCCCTGCAATGTTTAACGGACTATCTTCGGTTCGGGATATCGCAGAGGCC	1501
Db	1496	CGATCGGCGGTGCGAGCTGTTCAACCGTACTTTATTTGGATGGGATATTGACGCGGTA	1555
Qy	1502	AGAAGAAGATCTGGCCGACGTACAAGACGAATGGAAGCGGTACAGCTATACGATACACAGT	1561
Db	1556	CAACTCAGATTTGGCCCGAATATAGCACCAACGGGACATCGCAGGCTACACACGAGT	1615
Qy	1562	CGTCCGTGCGGTGCGCTACGCGTACACGATGTCAGT	1599
Db	1616	CGTTCGTGCCACTGTCCACGCGGTTCGAGTACCGGTCAAT	1653

```

RESULT 4
AX534814
LOCUS          AX534814            3150 bp    DNA             linear     PAT 22-NOV-2002
DEFINITION    Sequence 52 from Patent WO02068623.
ACCESSION     AX534814
VERSION       AX534814.1   GI:25261343
KEYWORDS
SOURCE        Aspergillus niger
ORGANISM      Aspergillus niger
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE     1
AUTHORS       Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock, A.,
               Kimpel, E., Kludbauer, S., Wagner, C., Fritz, A., gustet Von, W.,
               Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S.,
               Kemmer, W., Tan, P., Stiebler, J. and Albang, R.
TITLE         Novel genes encoding novel proteolytic enzymes
JOURNAL       Patent: WO 02086623-A 52 06-SEP-2002;
DSM N.V. (NL) Location/Qualifiers
               1. .3150
                  /organism="Aspergillus niger"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:5061"
BASE COUNT    801 a 778 c 738 g 833 t
ORIGIN

```

Query Match	19.5%;	Score 313.6;	DB 6;	Length 3150;
Best Local Similarity	72.5%;	Pred. No. 5.6e-70;		
Matches	406;	Conservative 0;	Mismatches 154;	Indels 0; Gaps 0;
QY	1037	ATGACCCGACTCGGCCAAGTTATTACAAACAATTTCTGCAAAAGACTCTGTCAATGGACG	1096	
DB	2194	AGGACCCGACCGCCCGTCTCTTTGTGTACTACCTCAAGAAAGACTCAGTCATGGATTG	2253	
QY	1097	CTATCGGGCTCAACATCACTACACCCAGTCCTCAATAATGACGCTTACTACGTTTTCACG	1156	
DB	2254	CTATCGGGCTGGACATTAACTACACCGAGTCCAGCGGCGAATATATTGCATTCCACG	2313	
QY	1157	AAACAGCGCACTTTGCTGGCCCAACTTCATCGAAGACCTCAGGAGATCCTTGCTTCG	1216	
DB	2314	AGACCGCGCACTTTGTATGSCCGAATTCATTACGACCTCGAAGATCCTCCACTCC	2373	
QY	1217	CGGTGGTGTCTCCCTCATCTATGCGGACGCGGATTACATCTGCAACTGGTTCGGCGGTC	1276	
DB	2374	CGGTACCGGTGCGTTGATCTACGGCGATGCCGACTATATCTGTAACCTGGTTCGGCGGTC	2433	

Qy	1277	AGGCCGTTTCCTCGCTGCAGCACTACCTCCAAAGCGGCCAGTTCGGAAGCGCAGGGTACA	13356
Db	2434	AGGCATCTCACTCGCAGTTAACTACCCCATGACGCTCAGTTCGGTGCACGGGATACA	24933
Qy	1337	CGCCCTGAAAGTCAACGGGCTCCAGTATGGGAAACTCGGAGTATGGTAATTTCCTCT	13966
Db	2494	CACCCATGACAGTAGATGGGTCGAATACGCTGAGACTCGGAGTATGGCACTTTTCGT	25533
Qy	1397	TCACTCGGCTCTATGCGCAGGCCATGAAGTCCCATACTACCGGCCCATCGCCTCCCTGC	14566
Db	2554	TCACCCGGGTATATCAGGCTGGGCACGAGTTCCAIATATCAACCATCGCAGCTGTC	26133
Qy	1457	AATGTTTAAACCGGACATCTCTCGTTGGGATATCGCAGAGGGCCGAAGAAGATCGGC	1516
Db	2614	AGCTGTTCAACCGTACTTTATTTGGATGGATATTGCAGCGGTACAACTCAGATTGGC	26733
Qy	1517	CCAGCTACAGACGAATGGAAAGCTACAGCTACGCAATACAGTCTCGTGGCGGTGC	1576
Db	2674	CCGAATATAGCAACCAACGGGACATFCGAGGCTACACACAGGAGTCGTTCTGTGCCACTGT	27333
Qy	1577	CTACGGGTACACAGCATGCC	1596
Db	2734	CCAGCGGCTGGAGTACCACC	27533

RESULT 5	AX534872	1872 bp	DNA	linear	PAT 22-NOV-2002
LOCUS	AX534872	110	from Patent	WO02068623.	
DEFINITION	AX534872				
ACCESSION	AX534872.1	GI:25261464			
VERSION					
KEYWORDS	Aspergillus niger				
SOURCE	Aspergillus niger				
ORGANISM	Aspergillus niger				
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;				
REFERENCE	1				
AUTHORS	Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock, A., Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gusted Von, W., Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S., Kemmer, W., Tan, P., Stiebler, J. and Albang, R.				
TITLE	Novel genes encoding novel proteolytic enzymes				
JOURNAL	Patent: WO 02068623-A 110 06-SEP-2002;				
	DSM N.V. (NL)				
FEATURES	Location/Qualifiers				
source	1..1872				
	/organism="Aspergillus niger"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:5061"				
BASE COUNT	402 a	434 c	570 g	466 t	
ORIGIN					

Query Match.	8.4%;	Score 134.4;	DB 6;	Length 1872;
Best Local Similarity	48.1%;	Pred. No. 1.4e-23;		
Matches 644;	Conservative 0;	Mismatches 661;	Indels 33;	Gaps 8;
QY	186	CCATACCTCTTCGGTCTCTCGAAGCCAGACAT-----AACCCAGAAACTGCACCTAT	239	
Db	282	CANTACCTTTTCGGTACTTTCCTCCGCCATCACCACACAAGTATACATCCCACT	341	
QY	240	CACATTGTGGTTCGAATGGTGGCCCTCGAAGCGAATCTTTTGATCGGTCTCTCGAAGAGTT	299	
Db	342	CACCATCTGATGAACGGGGGGGCGGGGATCTCCATGATGGGTATTTTCAAGAGAA	401	
QY	300	GGSCCCTTGCCATGTCAATTGACTTWTGATGACTACATCAACCCCTAC-----TCGTGGAA	356	
Db	402	CGSGCCATGTACTGTCAATAGSACTCGAATTCACGGCTATAATCCCTGGTCGTGAA	461	
QY	357	CGAGGTCTCCAAATTACTATTCTGTGCCAGCCATTGGAGTCGGGTTTTCATATAGTGA	416	
Db	462	TGAGTACGTGGAATGTTGTATATGACGACGGGTGCAGACGGGAATTTAGTATATCATGT	521	
QY	417	TACGGTTGATGGGTCCATAACCTCTAACTGGGTGTCGAAAAATTCGAGCTTTGCAGG	476	

Db 522 GTTGGAGAAATGGGACGTTAGAT---TTGAAATGAGACGTTTGGTGGGACGTTGGCG- 576  
Qy  
Db 477 AGTTAGGCGCGGTACCAACCAATGATGACCACTCTGATGACATACATACATCTTGGCG 536  
Db 577 -AGTCAGATGTGATGGGACGTTGATGAGGACGGTTAATGGGGAAGGCGCTTGGGT 635  
Qy 537 AGAGGCGGCTGGGAGATCTGCAAGGATTCCTTAGTGACTACTAGCTGGGACTTAG 596  
Db 636 TGGTTGACAGTTTGGTGGTGAATCTCTGAATATGTTCTCTGTTGACGGGAATGG 695  
Qy 597 GGTGACGCTTAAGGACTTACGCTGATGGACGAGCTATGGAGGCACTATGTCCTGC 656  
Db 696 TGGTGGTGATGACAGGCTGATATGACGGAGTATATGGGAGGCTATGGGACGGTATGACCGG 755  
Qy 657 ATCTTCAATCAATTTTACAGCAGATGAGAGAAATGCCAAGGTAGTGT---TAATGG 713  
Db 756 ATACACGGCGCTCTTCAGAGATGATGAGAGGATGAGAGTGGGAGTGAAGCAACCG 815  
Qy 714 TGTTCAGCTTAATTTCAACTCTCTGGGAATATTAAACGGCATATGACGAGGCGATCCA 773  
Db 816 GAAGAAGATCCATTGGATAGCGGCGCAATTAATGAGTGGTGTGATTTACTCTGCA 875  
Qy 774 GGCCCTTACTACCTGAATTCGCTGTGAACAATACCTACGCTATCAAGGCTGCAACGA 833  
Db 876 GGTCCCTTGTCTCTGAGCAGCGGTATAACAATACGATATGGGATCGAGGGAATCAATCG 935  
Qy 834 GACGCTTACAACTACATGAATTTGCAACCAATGCCAAATGCTGTCAGGATTTGAT 893  
Db 936 CAGCCTTACGAGCGGCTATGATAGTTGAGCAAGCCTGGCGGTTGCAAGGATATGAT 995  
Qy 894 TTCCACCTGCAACAGACAACCGACCGCATAGCTGACTACCGCTCTGCGCGGAAGC 953  
Db 996 CATGAGTGTGCGATCTGCGGAGCTCGG---AGATCCCTCATCATCTGCGAGGAGC 1052  
Qy 954 CACCAACATGTGACGAGCAATGTGAGGGGCCATACCTACGCTTGTGCTGCTGCTG 1013  
Db 1053 GTCGGACTGTGTCGGGAGATCAAGAGCCTGTATACGAATACCTCCGGCGGAGGATA 1112  
Qy 1014 GTATGATATTCGATCCATATGATGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1073  
Db 1113 CTAGGATAGCGGATTTACGCGGATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1172  
Qy 1074 GGCAAGGACTCTGTATGACGCTATGCGGCTCAACATCAACTACACCCAGTCCCAAT 1133  
Db 1173 GAATCGCCATGGGTGCAAGGACCTTGGGCTCCCGGTGAACTATACCAATGCTGCT 1232  
Qy 1134 TGAGCTTACTACGCTTCCAGCAACAGGCGACTTTGTCTGCGCCCAACCTTCATCGAGA 1193  
Db 1233 GGCAGTGGGAACAGTTTCGCTCGAGGGCGGATATCCGGGAAATGATCCCGCGGAAT 1292  
Qy 1194 CCTGAGGAGA-----TCCCTGCTCTCCCGTGGCTGCTGCTGCTGCTGCTGCTGCT 1244  
Db 1293 GATCGGGGATATGGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1352  
Qy 1245 CGCGGATATCACTGCAACTGGTTCGCGGCTGAGCGGCTTTCCCTGCTGCGCAACTATC 1304  
Db 1353 CCGGAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1412  
Qy 1305 CCAAGCGCCGAGTTCGAGACGCGAGGTACACCGCCCTGAAAGTCAACGCGCTGAGTA 1364  
Db 1413 GGATCGGAGAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1472  
Qy 1365 TGGGG---AAACTCGGAGTATGTTAATTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1421  
Db 1473 TGGGGTCTAGTAGGAGTATGAGAACTTCTGTTACCGGCTGCTGCTGCTGCTGCTGCTGCT 1532  
Qy 1422 TGAAGTCCCACTACCAAGCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1481  
Db 1533 TGAGGTGCCATTTTATCAGCCGGAACCGGCTATGAGATTTTAAATCGCTGCTGCTGCTGCT 1592  
Qy 1482 TTGGGATATCCGAGGG 1499  
|||||

Db 1593 TTGGATATTCGACGGG 1610  
RESULT 6  
AX534866  
LOCUS AX534866 1665 bp DNA linear PAT 22-NOV-2002  
DEFINITION Sequence 104 from Patent WO02068623.  
ACCESSION AX534866  
VERSION AX534866.1 GI:25261452  
KEYWORDS Aspergillus niger  
SOURCE Aspergillus niger  
ORGANISM Aspergillus niger  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
REFERENCE 1  
AUTHORS Edens,L., dijk Van,A.A., Krubasik,P., Albermann,K., Stock,A.,  
Kimpel,E., Klugbauer,S., Wagner,C., Fritz,A., gustedt Von,W.,  
Heinrich,O., Maier,D., Spreafico,F., Folkers,G., Hopper,S.,  
Kemmer,W., Tan,P., Stiebler,J. and Albarg,R.  
Novel genes encoding novel proteolytic enzymes  
Patent: WO 02068623-A 104 06-SEP-2002;  
DSM N.V. (NL)  
FEATURES Location/Qualifiers  
source  
1..1665  
/organism="Aspergillus niger"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:5061"  
BASE COUNT 374 a 455 c 424 g 412 t  
ORIGIN  
Query Match 7.4%; Score 118.6; DB 6; Length 1665;  
Best Local Similarity 47.2%; Pred. NO. 1.7e-19;  
Matches 673; Conservative 0; Mismatches 649; Indels 105; Gaps 6;  
Qy 121 GCGCTGTCGAGACTACCGGGTGTCAAACTCTCTGATATGTCGACACCTCTCC 180  
Db 130 GGAATTTGCGAACTACCGGGTGTCCGATCTTAITCGGGTATGTACACCTTCCCGCC 189  
Qy 181 GAGTCCCATACCTCTCTGCTGCTGCGAAGCAGACATACCCAGAACTGACATC 240  
Db 190 GCTTCACACGCTCTTTGGTTTTTTCGAGCCGCAAGATCCAGCAATGCGGCTCTG 249  
Qy 241 ACATTGGTGTGATGGTGGCCCTGGAAGCGATTTTGTATCGGCTCTTTCGAGAGTTG 300  
Db 250 GCGATCTGCTCAATGCGGCTCGCGGTGCTGCTCATGGGCTCTTGAAGAATA 309  
Qy 301 GCGCTTGGCATGCAATTCGACTTTTGTATGAC---TACATCAACCCCTCACTCGTGAAC 357  
Db 310 GGTCTTTTCCATTCGATCAGATCCAGACACAGTCCATCTTCCTGATGATGTC 369  
Qy 358 GAGGTCTCCAAATTTACTATTCTGTCGCCAGGCAATTTGGGAGTGGGCTTTTCATATAGT 417  
Db 370 AATGAAGTCAATCTCTATCTCTTTCGACGAGCAACTCAAGTCTGCTTCTCATAGATGTC 429  
Qy 418 ACGGTGTGATGGTTCATTAACCTGTAACCTGTAACCTGCGGCTGCGAATAATTCAGCTTTC 477  
Db 430 CCAACAAATGGCATTTCAGACGTA-----ATGGGACTTCGATTCGCGCT 454  
Qy 478 GTTCAGGCGCGGTACCCCAACCATTCGATGCCACTCTGATGATCACTACCAATCTTCGCCGA 537  
Db 455 -----ATGGGACTTCGATTCGCGCT 474  
Qy 538 GAGGCGGCTTGGGAGATCTCGAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGG 597  
Db 475 CACGCTCTATGGCAATTTCCGCGCAACCTGTTTTCGAGTTCGCCACACTACAAGCAAC 534  
Qy 598 GTGAGTCTAAGGACTTCAGTCTATGCGAGGAGACTATGGAGGCACTATGGTCTCTGCA 657  
Db 535 GATGATCTG-----TCAGTCTCTGGCTGAAAGTTACGGAGGCAATATGTTCCAGGC 588  
Qy 658 TCTTCAATCATTTTACGAGCAATGAGAGAAATTCGCAACGCTAGTGTATGATGCTGTT 717  
Db 589 ATCTTTCGCTTCTCCACAGCAGATGACAAATTCGAGAGGCGACTGCAAGACGCT 648

QY 718 -----CAGCTTAATTTCACTCTCTGGGAATTAATTAACGGCATCATCGACGAGCGATC 771  
Db 649 GCACAGTATTGCACTCTCGACACGCTTGGCATTTGACGCGCTTATGATGATGATGATC 708  
QY 772 CAGGCCCTTACTACCCIGATTCGCTGTGAACAAATACCTACGCTATCAAGCGTGTCAAC 831  
Db 709 CAAGAAGAGGCTTACATTTACTTGTGCCATACAAATA--ACGTAAGCGTCGCCCTTCTTCAT 766  
QY 832 GAGACCTCTACACTACATGAAGTTTCCCAACCAATGCCAAAGTGTGTCAGCAATTTG 891  
Db 767 TCACTCGCAGGCTTTCGCGATCAGGCCCTCGCGCTCGAAGCGGCTTTGAAGAAGCGG 826  
QY 892 ATTTCCACCTCGCAACAGACAAACCGCACGCAATTAAGTACTAGCTACTAGCGCTCTGCGCGAA 951  
Db 827 ATTCCGGCTTGC-----TCACTCAGGGAAGATATCTCTGAAATTTGGGA 873  
QY 952 GCCACCAACATGTCAGGACATATGTTGAGGGGCCATACACTACGCGCTTTGCGTGTGGT 1011  
Db 874 GGCCTTGCACTAGAAATGGGAGATGGGCCATACCTACTACTACCACTTCAATCGCGGG 933  
QY 1012 GTGTATGATATCGGCATCATATGATGACCGCACTCGCCAACTTATACAAATTT 1071  
Db 934 TGGTACGACATCGCCATCTAAGAACGCCATTCCTGCAAGACATGCTCGGAT 993  
QY 1072 CTGCAAGGACTCTGTCTATGAGGCTATCGCGGTCAACATCAACTACACCCAGTCCAAT 1131  
Db 994 TTGACGAGGAGTCCGCTTCTGCGCTTGGGGTACCACTATTTCACTCTGCTCTCG 1053  
QY 1132 AATGACGCTTACTACGCTTTCAGCAACAAAGGCGACTTTGCTGTGGCCCACTTCAATCGAA 1191  
Db 1054 AGTGGCGTGGCTACACAGATTTCATAAAACCTTTGATTCCTCCACGCGGCTTCCIGAT 1113  
QY 1192 GACCTCGAGGAGATCCTTGTCTCTCCGCTGGGTGCTCCCTCTCGCTGGAATCTACTCCCAAGCC 1311  
Db 1114 GCATTTGGCTACCTCTCGACATGGTGTAAAGTACACATGATGTACGGAGATCTGTGAT 1173  
QY 1252 TACATCTGCACTGGTTCGCGGTGACGGCGTTTCCCTCTGCTGGAATCTACTCCCAAGCC 1311  
Db 1174 TAGCGCTGCAATTTGGTTCGGGGCGAAAGCCAGCGCTTCGAGTTCGGTATCCCGTATC 1233  
QY 1312 GCCAGTTCGAGCGCAGGTACACGCCCTGAAAGTCAACGGCGTCGAGTATGGGAA 1371  
Db 1234 ACCGAATTTGCCACACGGGATACCTCCCACTCTTACGCCCGCGGATCAGCGGCATG 1293  
QY 1372 ACTCGGAGTATGTTAAATTTCTCTTCACTCGGCTCTATGACGACGCCATGAGTCCCA 1431  
Db 1294 ACCGCCAGTGGCAACTACAGCTTCACTCGCGCTTCCCAAGCGGCGATGAGGTCCCC 1353  
QY 1432 TACTACAGCCCATCGCTCCCTGCAATTTGTTTAAOCCGACTATCTTGGTGGGATATC 1491  
Db 1354 TCTACAGCGCTGTCGGCGATGATGATCTTCATCGCGGCGACATTCACAAAGATAIC 1413  
QY 1492 GCAGAGGGCCAGAGAAGATCTGCGCCCGACGCTACAGCAAGATGGAAC 1538  
Db 1414 CCTACTGGCTCTTGGCTGTGTGATGAGAAATCCAGTCCGTTGGACC 1460

RESULT 7  
AX534824  
LOCUS AX534824  
DEFINITION Sequence 62 from Patent WO02068623.  
ACCESSION AX534824  
VERSION AX534824.1 GI:25261366  
KEYWORDS Aspergillus niger  
SOURCE Aspergillus niger  
ORGANISM Aspergillus niger  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
REFERENCE 1  
AUTHORS Edens,L., dijk Van,A.A., Krubasik,P., Albermann,K., Stock,A.,  
Kimpel,E., Klugbauer,S., Wagner,C., Fritz,A., gustedt Von,W.,  
Heinrich,O., Maier,D., Spreafico,F., Folkers,U., Hopper,S.,

Kemmer,W., Tan,P., Stiebler,J. and Albang,R.  
Novel genes encoding novel proteolytic enzymes  
Patent: WO 02068623-A 62 06-SEP-2002;  
DSM N.V. (NL)  
FEATURES  
source Location/Qualifiers  
1..1581  
/organism="Aspergillus niger"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:5061"  
BASE COUNT 343 a 473 c 423 g 342 t  
ORIGIN  
Query Match 7.3%; Score 116.8; DB 6; Length 1581;  
Best Local Similarity 52.5%; Pred.No.5.le-19;  
Matches 334; Conservative 0; Mismatches 287; Indels 15; Gaps 3;  
QY 914 ACCGACCGCATTTAGCTAGCTACGCGCTTCGCGGGAAGCCACCAACATGTCGAGGACA 973  
Db 812 ACTGGCGCGCGGAGGATCCAGGAGATCTGAGCACTGCGGACGATTTTTCGCGCAACG 871  
QY 974 ATGTTGAGGGCCATACACGCTTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1033  
Db 872 AGGTGAAAAGCTCTAGACATTTACTTCGGTTCGGGATGAGTATGACTTTCGTGAATCA 931  
QY 1034 ATGAAGACCGACTCCGCCAAGTTTATACACAAATTTCTGCGAAAGAGCTGTGTATGG 1093  
Db 932 CTCGGACCGCTTCCCTTACGAGTTCACGTTGACTACCTGAACAAAGCGTCCGTGCAG 991  
QY 1094 ACGTATCGGGCTCAACATCACTACCCAGTCCCAATTAAGTCTACTACGCTTTC 1153  
Db 992 CCGCATCGGGCATACATCAATTTACAGGAGAGACAAACGCTGTGGACTCGCCCTTT 1051  
QY 1154 AGCAACAGCGGACTTTGTCTGGGCC---AATTTCAFGAAGAGCTCGAGAGATCCTTG 1210  
Db 1052 CGTCCAGCGGTGACGAGCGGGGACTCATGAACACCATCCAGGATGTGGCAAGCTGTCA 1111  
QY 1211 CTCFCCCGGTGCTCTCCCTCATCTATGCGGACGCGATATACATCTGCAACTGTTCG 1270  
Db 1112 AACAGGTGTACCGGTGTATGACCGCGGGATGCGGACTATACTGCAACTGGCTGG 1171  
QY 1271 CGGCTAGCGCGTTTCCTCGTCACTACCTCCCAAGCGCGCTTCGGAAGCGCAG 1330  
Db 1172 GTGGGAAGCGGTG-----CGTTGAGGTCAAGCGCGCACTTCAGTAGTCGG 1222  
QY 1331 GTTACAGCGCC---CTGAAAGTCAACGCGTTCGAGTATGGGGAACCTCGGAGTATGTA 1387  
Db 1223 GTTACACCAATTTGCACCTCGGATGAGTACACACGCGGAGTTCGCGGCGGCG 1282  
QY 1388 ATTTCTCTTCACTCGGCTATGAGGACGCGCATGAAGTCCCATACCTACAGCCCATCG 1447  
Db 1283 AATTGCTTTGCGAGTGTATGAGAGTGGATGAGTTCCTCTCTATCAACCCCTTGC 1342  
QY 1448 CTTCCCTGCAATTTTAAACCGGACTATCTTCGTTGGGATATCGCAGAGGCGCAAGA 1507  
Db 1343 TTGCGCTGGAGATTTTGAAGCGCTCATTTGGGCAAGGATGTGGCCACGGGAAGATTC 1402  
QY 1508 AGATCTGGCCCGCTACAGCAAGCAAGTGAACGGCTA 1543  
Db 1403 CCATCTCGTGGAGTTTACAGACGTTGGGCGACGCCA 1438  
RESULT 8  
AX534767  
LOCUS AX534767  
DEFINITION Sequence 5 from Patent WO02068623.  
ACCESSION AX534767  
VERSION AX534767.1 GI:25261241  
KEYWORDS Aspergillus niger  
SOURCE Aspergillus niger  
ORGANISM Aspergillus niger  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
REFERENCE 1  
AUTHORS

AUTHORS	Edens, L., dijk Van, A. A., Krubasik, P., Alberman, K., Stock, A., Kimpel, B., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W., Heimerl, O., Maier, D., Spreafico, F., Folkers, O., Hopper, S., Kemmer, W., Tan, P., Stiebler, J., and Albang, R.									
TITLE	Novel genes encoding novel proteolytic enzymes									
JOURNAL	Patent: WO 02068623-A 5 06-SEP-2002;									
FEATURES	DSM N.V. (NL)									
source	Location/Qualifiers									
	1. .2940									
	/organism="Aspergillus niger"									
	/mol_type="genomic DNA"									
	/db_xref="taxon:5061"									
BASE COUNT	691 a 818 C 732 g 699 t									
ORIGIN										
	Query Match 7.3%; Score 116.8; DB 6; Length 2940;									
	Best Local Similarity 52.5%; Pred. No. 5.4e-19;									
	Matches 334; Conservative 0; Mismatches 287; Indels 15; Gaps 3;									
Qy	914	ACCGCACCGCATTTAGCTGACTAGCCCTTCGCGCCGAAGCCACCAACATGTGCAGGACA	973							
Db	1492	ACTGCGCGCCCGCGAGCATCGGAGAGATCTGCAGCACTGCCGACGATTTTGGCCCAACG	1551							
Qy	974	ATGTTGAGGGGGCCATACTAGCCCTTCGCTGCTCGTGGTGTATGATATTCGCATCCAT	1033							
Db	1552	AGGTCGAAACGCTACGACATTTACTCCGTCGGATGAGTATGACTTTCGFIGAACTCA	1611							
Qy	1034	ATGATGACCCGACTCCGCCAAGTATTATCAACAATAATTCGGCAAGAGACTCTGTATGS	1093							
Db	1612	CTCCGACCCGTTTCCCTTAGAGTTCTACGTTTACTACCTGAACAAGCGTCGTCGAGG	1671							
Qy	1094	ACGCTATCGCGCTCAACATCACTACACCCAGTCCCAATATGACGCTACTACGCTTTCC	1153							
Db	1672	CGGCCATTCGGCGATACATCAATACACGAGAGCACACGCTGTTGGACTCGCCTTTT	1731							
Qy	1154	AGCAACAGCGACTTTGTCTGGCCC---AACTTCATCGAAGACCTCGAGGAGATCCTTG	1210							
Db	1732	CGTCCACGGTGCAGCGGGCGACTCATGAACACCACTCCAGGATGTGGGCAAGCTGTCA	1791							
Qy	1211	CTCTCCCGGTGGTGTCTCCCTCATCTATGCGGACGCCGATATCATCTGCAACTGGTTCG	1270							
Db	1792	AACAGSGTGTACCGGTGGTCTATGACCGCGGGATGCCACTAACTGCACTGGCTGG	1851							
Qy	1271	CGGCTCAGGCGGTTTCCCTCGCTCGCAACTACTCCCAAGCGGCCAGTCCGAGGCGCAG	1330							
Db	1852	GTGGGGAAGCGGT-----CGTTGCAAGTCAAGCGGCCCACTTCAGTAGTGGG	1902							
Qy	1331	GGTACAGCCCC---CTGAAAGTCAACGGCGTGAGTAGTGGGGAACATCCGCGATGTGTA	1387							
Db	1903	GTATACCAACATTTGTACCTCGATGGAGTGACACACGGCCAGTGTGCGCAGCGGGGGC	1962							
Qy	1388	ATTTCCTCTTCACTCCGGTCTATGAGCAGCGCCATGAAGTCCCATATCAACGCCCATCG	1447							
Db	1963	AAATTGGCTTTGTCCGAGTGTATGAGAGTGACATGAGTTCCTTTCTATCAACCCITGC	2022							
Qy	1448	CCTCCCTGCAATTTGTTAAACCGGACTATCTTCGGTTGGGATATCGCAGAGCGGCACA	1507							
Db	2023	TTGGCTGGGAGTGTTTACGGCCGTCATTTGGCGCAAGATGTGGCGACGGGAAGATTC	2082							
Qy	1508	AGATCTGCCCGCACTACAAGACGAAATGAACGGCTA	1543							
Db	2083	CCATCTCGTTCGAGTTTACAGACGGTGGCGACGCCA	2118							

RESULT	9
AX534809	
LOCUS	AX534809
DEFINITION	Sequence 47 from Patent WO02068623.
ACCESSION	AX534809
VERSION	AX534809.1
KEYWORDS	. GI:25261332
SOURCE	.
ORGANISM	Aspergillus niger
	Aspergillus niger
	linear
	3080 bp DNA
	PAT 22-NOV-2002

[illegible]

RESULT 10	
AX534815	
LOCUS	
DEFINITION	Sequence 53 from Patent WO02068623.
ACCESSION	AX534815
VERSION	AX534815.1 GI:25261346
KEYWORDS	.
SOURCE	Aspergillus niger
ORGANISM	Aspergillus niger
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;



Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

**REFERENCE**  
 1. Edwards, L., Trichomycetaceae, mitosporig Trichomycetaceae; Aspergillus  
**AUTHORS**  
 Edens, L., dijk Van, A., Kribasik, P., Albermann, K., Stock, A.,  
 Klumpe, E., Klubbauer, S., Wagner, C., Fritz, A., gustedt, V., W.,  
 Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S.,  
 Kemmer, W., Tan, P., Stiebler, J. and Albarg, R.  
**TITLE**  
 Novel genes encoding novel proteolytic enzymes  
**JOURNAL**  
 Patent: WO 02068623-A 53 06-SEP-2002;

```

DSM N.V. (NL)
Location/Qualifiers
1. .3221
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon.5061"

```

```

BASE COUNT      739 a      773 c      869 g      840 t
ORIGIN

Query Match      6.7%;      Score 107.8;      DB 6;      Length 3221;
Best Local Similarity 48.5%;      Pred. No. 1.2e-16;
Matches 449; Conservative 0; Mismatches 437; Indels 39; Gaps 4;

```

QY	614	TCAGTCTATGGACGGACAGCTATGGACGGCACATATGGTCTCGCATTCCTCAATCATATTTT	673
Db	1522	TGAGTATATGGACGGAGCTATATATGGGGGACGGTATGGACCGGACATACACGGCGGTCTTTC	1581

[illegible]

731	ACTCTCTGGGAATTATTACGGCAI	CATCACGAGGGATCCAGGCCCTT	TACTACCTG	790
QY	- - - - -	- - - - -	- - - - -	
1642	ATACGCTGGCAATATCAATGGTGT	GTGATTTACTCTGCAGGTCCTT	CGTTCCTG	1701
Db	- - - - -	- - - - -	- - - - -	

791 AATTCGCTGGAACANTACCTACGCTATCAAGCTGTCAACGAGACGCTCACTAACA 850  
 1702 AGCAGCGGTATACANTACGCTATGGATCGAGGAATCAATCGCACCTCTACGACCGG 1761

851 TGAAGTTTGCCACCAAAATGCCAAATGTTGCCAGGATTTGATTTCCACCIG----- 902 QY  
1762 CTAATGATAGTTGGAGCAACCTGGCGGTGCAGGATATGATCAATCGATGTGGCGCATC 1821 Db

QY 903 -----CAAACAGACAACCGCAGCGATTAGTGTACTAGCGCCTCTGGC 946  
 1822 CTGGGAGCTGGGAGTACCCCTCATGTATGGCCACAAATGAGCGTGAATAGCACTCCG 1861

947 CGGAAGCCACCAACATGTGCAGGAGCAATGTTGAGGGGCCAIXACTAGCCTTTGCTGGTC 1006

1882 AGGAGCGTCGGACTACTGTTCGGGAGATCAAGACCTGTATCAATACCACCTCGGGC 1941

**Qy** 1007 GGGTGTGATGATTCCGTCATCCATATGATGACCGGAGTCGCCAAGTTATTACAACA 1066  
| | | | | | | | | | | | | | | |  
**Dβ** 1942 GAGGATACTACGACATAGCGCATTTCCAGCGCGGATGCAGGCTCTCGTGCCCTACTTCGTGCG 2001

Qy	1067	AATTTCTGCAAGGACTCTGT 	CATGGAGCTATCGGCCTCAACATCAACTTCAACCAGT	1126
Db	2002	GGTCTTGTAATGCCCATCGG 	TGCAAAAGGCACCTGGGGTCGCGGTGAACATAACCATGT	2061

[illegible]

Qy  
1178 CCAACTTCATCGAAGACCTCGAGGAGATCCTTGCTCCCGGTGGTGTCCCTCAICT 1237

Dd  
2122 GCGGAATGATCGGGGATATTGGATACTTGTTGACTCCGGTGTCAAGGTGGCATGGTAT 2181

QY  
1238 ATGGGACCGGATTACATCTGCAACTGTTCGGCGTTCAGGCCGTTCCICIGCTCCGA 1297

Dδ  
2182 ATGGGACCGGACIATGTTTCCTGGCGCGGGAGAATATGCAGCTCCTGSGTG 2241

Qy	1298	ACTACTCCCAAGCGGCCAGTTCGGAAGCAGGGGTACAGGCCCTGARAGTCAACGGCG	1357
Db	2242	AGTAGGAGATCGGAGAAGTTCCGTGCTGGGTATGCCGAAAGTGACACAAGATCAT	2301

QY	1358	TCGAGTATGGG---AAACTCGCAGATGGTAATTTCTCTTCACTCGGCTCATGAGG	1414
Db	2302	CCTACGTTGGGGTCTACIANGGCAGATATGSGAACTTCTCGTTCACGGGTGCTTTCAGG	2361
QY	1415	CAGGCCATGAAGTCCCATCTACTACACGCCCATCGCTCCCTGCAATTGTTAAACGGACTA	1474
Db	2362	CGGCGCATGAGTGCCATTTTATCAGCCCGAAACGGCGTATGAGATTTTAAATCGCGCTC	2421
QY	1475	TCCTCGGTTGGGATATCGCAGAGG	1499
Db	2422	AGTTAATTTGGGATATTGCGACGGG	2446

RESULT 11				
A75535	A75535	2503 bp	DNA	linear
LOCUS	Sequence 2 from Patent WO9400579.			
DEFINITION	A75535			
ACCESSION	A75535			
VERSION	A75535.1	GI:6065609		
				PAT 15-OCT-1999

KEYWORDS  
SOURCE  
ORGANISM  
unidentified  
unidentified  
unclassified  
1 (bases 1 to 2503)  
REFERENCE  
Fleer R. and Fournier A.  
MODIFIED KLOVERMYCES YEASTS, THEIR PREPARATION AND USE  
Patent: WO 9400579-A 2 06-JAN-1994;  
JOURNAL  
RHONE POULENC RORER SA (FR); FLEER REINHARD (FR)  
Location/Qualifiers  
1. .2503  
FEATURES  
source

```

gene
/organism="unidentified"
/mol_type="genomic DNA"
/strain="Kluyveromyces lactis"
/db_xref="taxon:32644"
387..1862
/gene="Kl.PRC1"

CDS
387..1862
/gene="Kl.PRC1"
/codon_start=1
/product="Gene de la protease C de K.lactis"
/protein_id="CAB58521.1"
/db_xref="CCT:6065610"

```

/translation="MYSIKELLSYGLWSVTIAISLNAVVDLSFNSFGNNNIEDHE  
 TANYTFQVSSNDDASLRPLDKPSGLVDYTKWMSGLDYQDSKHFWFEPES  
 RENDPNDVPLWINGSGSGVFLFELFQSPSSIGADLPYINPYSWNSNASVFLQD  
 PVGSGYSGDSKSTLDDAAKDYIFLDLFFERFLRNNDPHISGESVAGHLPKIA  
 HETAAVHARDSSVSLSLNGFETDPLTYOYVPMWACGEGYPAVLEPCLDMNR  
 NLPLCLSLVDRCKYKSHVSCVLAADRYCEQITGVYKSGRNPYDIRSKCEADDSGA  
 CYQREIYISDYLNQEEVRALGTDVSSFGSCSDVGIGFATGDPGSPHQVAAELLD  
 QDNLVLIAGDKDGYICNWLGNLAWTEKLEWRNYYKKQVLRTKTSSEETDGETIGKS  
 YGPLTYLRIYDAGHWYHPDQENSLWNSWICNIAKRSRI"  
 703 a 517 q 808 t  
 BASE COUNT 475 c

Query Match	5.4%	Score 86.4;	DB 6;	Length 2503;
Best Local Similarity	57.4%	Pred. No. 4e-11;		
Matches 156: Conservative		0: Mismatches 116:	Indels 0:	Gaps 0:

QY	145	GTCAATCTCTACTCTGGATATGTCACACCTCTCCCGAGTCCCATACCTTCTTCTCGTTC	204
Db	624	GTGAACAATGGTGGGATATTTAGATTACAGAGCTCAAAACACTTTTATTGGTTT	683
QY	205	TTCGAAGCCAGACATAACCCAGAACTGCACCTATCACATTGTGGTTGAATGGTGCCCT	264
Db	684	TTTGAGTCTAGAATGACCCAGAGATGACCCAGTAGTACTATGGTTAAAGGTGTGCTC	743
QY	265	GSAGCGAATCTTTGATCGCTCTCTCGAAGAGTGGGCCCTTGCCTATGCAATTCGAT	324
Db	744	GGCTGTGCCCTTTTCGTGGCTCTTTCTTIGANTGGACCTTCTCTATAGAGCTGAT	803
QY	325	TTTGAGCTACATCAACCTCTACTCGTGAACGAGTCTCCAAATTTACTATTCCTGCC	384
Db	804	TTGAACCCATTTATACCCCIACCTCTGGAATTCGAATTCGAAGCTCTCTGTGATATCCTAGAT	863

QY 385 CAGCCATTGGAGTCGGCTTTTCATATAGTGA 416  
 ||||| | | | | | | | | | | | | | | | |  
 Db 864 CAGCCATTGGAGTCGGCTTTTCATACGGTGA 895

RESULT 12  
 I70282  
 LOCUS I70282 2503 bp DNA linear PAT 03-APR-1998  
 DEFINITION Sequence 3 from patent US 5679544.  
 ACCESSION I70282  
 VERSION I70282.1 GI:3006417  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE  
 1 (bases 1 to 2503)  
 AUTHORS Fleer, R., Fournier, A. and Yeh, P.  
 TITLE Modified Kluyveromyces yeasts, their preparation and use  
 JOURNAL Patent: US 5679544-A 3 21-OCT-1997;  
 FEATURES  
 Location/Qualifiers  
 source 1..2503  
 /organism="unknown"  
 BASE COUNT 703 a 475 c 517 g 808 t  
 ORIGIN

Query Match 5.4%; Score 86.4; DB 6; Length 2503;  
 Best Local Similarity 57.4%; Pred. No. 4e-11;  
 Matches 156; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 145 GTCAAATCTTACTCTGGATATGCGACACCTCTCCGGAGTCCCATACCTTCTTCTGTTTC 204  
 ||||| | | | | | | | | | | | | | | | |  
 Db 624 GTGAACAATGGTCGGGATATTAGATTACCAGGACTCAAAACACTTCTTTTATGGTTT 683

QY 205 TTGCAAGCAGACATACCCAGAACTGCACCTATACATGTGTGGTTGAATGGTGGCCCT 264  
 ||||| | | | | | | | | | | | | | | | |  
 Db 684 TTGAGCTAGAAATACCCAGAGAATACCCAGTATAGTATGTTTAAACGGTGGTCT 743

QY 265 GGAAGCGATTCTTTGATCGTCTCTCGAAGAGTGGGCCCTGCCATGTCATTCGACT 324  
 ||||| | | | | | | | | | | | | | | | |  
 Db 744 GGCTGTCTCTCTCGTCTGCTCTTTCTTTGAATGGGACCTCTCTTATAGGAGCTGAT 803

QY 325 TTGATGACTATACATCAACCTCTCTGTTGAAGAGTCTCCAAATTTTACTTATTCCTGTC 384  
 ||||| | | | | | | | | | | | | | | | |  
 Db 804 TTGAACCCATTTATTAACCCCTCTCTGGAATTCACACCTCTCTGATATTCCTAGAT 863

QY 385 CAGCCATTGGAGTCGGCTTTTCATATAGTGA 416  
 ||||| | | | | | | | | | | | | | | | |  
 Db 864 CAGCCATTGGAGTCGGCTTTTCATACGGTGA 895

RESULT 13  
 I33983  
 LOCUS I33983 2002 bp DNA linear PAT 06-FEB-1997  
 DEFINITION Sequence 3 from patent US 5594119.  
 ACCESSION I33983  
 VERSION I33983.1 GI:1824774  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE  
 1 (bases 1 to 2002)  
 AUTHORS Yaver, D.S. and Thompson, S.A.  
 TITLE Gene encoding carboxypeptidase of aspergillus niger  
 JOURNAL Patent: US 5594119-A 3 14-JAN-1997;  
 FEATURES  
 Location/Qualifiers  
 source 1..2002  
 /organism="unknown"  
 BASE COUNT 416 a 591 c 505 g 490 t  
 ORIGIN

Query Match 5.2%; Score 84; DB 6; Length 2002;  
 Best Local Similarity 55.5%; Pred. No. 1.6e-10;

Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 169 GACACCTCTCCGAGTCCCATACCTTCTTCTGGTCTTTCGAAGCCAGACATAACCCAGAA 228  
 ||||| | | | | | | | | | | | | | | | |  
 Db 617 GACAACGAGAACGACAAGCACTGTCTACTGTCTTCGAGTCTCGCAATGACCCGAG 676

QY 229 ACTGCACCTATCATCATTTGGTTTGAATGGTGGCCCTGGAAGCGATTTTGTATGGGTCTC 288  
 ||||| | | | | | | | | | | | | | | | |  
 Db 677 AATGACCCCTGTCTGTGGTGAACGGTGGCCCTGGATGCTTCTTCCCTCACCGGTCTT 736

QY 289 TTCGAAGAGTTGGCCCTTGCCTGCAATTCGACATTTTGTGACATACATCAACCCCTAC 348  
 ||||| | | | | | | | | | | | | | | | |  
 Db 737 TTTATGAGCTCGGCCCTTAGCAGCATCAACAAGAATCCAGCGGTCTACAACGACTAC 796

QY 349 TCGTGGAAAGAGGTCTCCAAATTTACTATTCTCTGCCAGCATTTGGGAGTGGGTCTTCA 408  
 ||||| | | | | | | | | | | | | | | | |  
 Db 797 GCTTGGAACTCCAAACGGTCCGTGATCTTCTTGACCAGCTGTCAACGTCGGTACTCT 856

QY 409 TATAGTATACGGTTGATGGTCCATTAACCCCTGTAACCTGGGTGTCGCA 458  
 ||||| | | | | | | | | | | | | | | | |  
 Db 857 TACAGCAACTCTGCTGTACGAGACACCGTTGCTGCTGGCAAGGACGTCTA 906

RESULT 14  
 I74375  
 LOCUS I74375 2002 bp DNA linear PAT 03-APR-1998  
 DEFINITION Sequence 3 from patent US 5688663.  
 ACCESSION I74375  
 VERSION I74375.1 GI:3010516  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE  
 1 (bases 1 to 2002)  
 AUTHORS Yaver, D.S. and Thompson, S. Ann.  
 TITLE Gene encoding carboxypeptidase of Aspergillus niger  
 JOURNAL Patent: US 5688663-A 3 18-NOV-1997;  
 FEATURES  
 Location/Qualifiers  
 source 1..2002  
 /organism="unknown"  
 BASE COUNT 416 a 591 c 505 g 490 t  
 ORIGIN

Query Match 5.2%; Score 84; DB 6; Length 2002;  
 Best Local Similarity 55.5%; Pred. No. 1.6e-10;  
 Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 169 GACACCTCTCCGAGTCCCATACCTTCTTCTGGTCTTTCGAAGCCAGACATAACCCAGAA 228  
 ||||| | | | | | | | | | | | | | | | |  
 Db 617 GACAACGAGAACGACAAGCACTGTCTACTGTCTTCGAGTCTCGCAATGACCCGAG 676

QY 229 ACTGCACCTATCATCATTTGGTTTGAATGGTGGCCCTGGAAGCGATTTTGTATGGGTCTC 288  
 ||||| | | | | | | | | | | | | | | | |  
 Db 677 AATGACCCCTGTCTGTGGTGAACGGTGGCCCTGGATGCTTCTTCCCTCACCGGTCTT 736

QY 289 TTCGAAGAGTTGGCCCTTGCCTGCAATTCGACATTTTGTGACATACATCAACCCCTAC 348  
 ||||| | | | | | | | | | | | | | | | |  
 Db 737 TTTATGAGCTCGGCCCTTAGCAGCATCAACAAGAATCCAGCGGTCTACAACGACTAC 796

QY 349 TCGTGGAAAGAGGTCTCCAAATTTACTATTCTCTGCCAGCATTTGGGAGTGGGTCTTCA 408  
 ||||| | | | | | | | | | | | | | | | |  
 Db 797 GCTTGGAACTCCAAACGGTCCGTGATCTTCTTGACCAGCTGTCAACGTCGGTACTCT 856

QY 409 TATAGTATACGGTTGATGGTCCATTAACCCCTGTAACCTGGGTGTCGCA 458  
 ||||| | | | | | | | | | | | | | | | |  
 Db 857 TACAGCAACTCTGCTGTACGAGACACCGTTGCTGCTGGCAAGGACGTCTA 906

RESULT 15  
 I77239  
 LOCUS I77239 2002 bp DNA linear PAT 03-APR-1998  
 DEFINITION Sequence 3 from patent US 5693510.  
 ACCESSION I77239

Search completed: September 16, 2003, 18:04:51  
Job time : 7747 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 14:42:53 : Search time 148 Seconds  
(without alignments)  
4795.568 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662

Perfect score: 1608

Sequence: 1 cttccaggaagtacacggc.....gcattgtccagtgttggtatg 1608

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 559978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCITUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1608	100.0	1662	3	US-08-943-714-1
2	86.4	5.4	2503	1	US-09-640-305-3
3	86.4	5.4	2503	1	US-08-360-673-3
4	84	5.2	2002	1	US-08-309-341-3
5	84	5.2	2002	1	US-08-608-267-3
6	84	5.2	2002	1	US-08-608-452-3
7	84	5.2	2002	1	US-08-608-224-3
8	84	5.2	2002	2	US-08-967-149-3
9	76	4.7	2068	1	US-08-309-341-1
10	76	4.7	2068	1	US-08-608-267-1
11	76	4.7	2068	1	US-08-608-452-1
12	76	4.7	2068	1	US-08-608-224-1
13	76	4.7	2068	2	US-08-967-149-1
14	66.2	4.1	2632	2	US-08-899-324-32
15	66.2	4.1	2632	3	US-08-329-892B-32
16	56.4	3.5	1814	4	US-09-702-705-319
17	56.4	3.5	1814	4	US-09-736-457-319
18	47.6	3.0	390	3	US-09-197-649-7
19	47.6	3.0	1551	2	US-08-828-488-4
20	47.6	3.0	1551	4	US-09-299-689A-4
21	47.2	2.9	1670	2	US-08-828-488-2
22	47.2	2.9	1670	4	US-09-299-689A-2
23	39.8	2.5	357	4	US-09-280-116-131
24	38.6	2.4	447	4	US-09-252-991A-11967
25	38.6	2.4	630	4	US-09-252-991A-8154
26	38.6	2.4	774	4	US-09-252-991A-11893
27	38.6	2.4	777	4	US-09-252-991A-8222

28	38.6	2.4	1083	4	US-09-252-991A-11930
29	38.6	2.4	1239	4	US-09-252-991A-8276
30	38.6	2.4	1434	4	US-09-252-991A-8046
31	38.2	2.4	918	3	US-09-105-390-35
32	38.2	2.4	1008	3	US-09-105-390-51
33	38.2	2.4	1949	3	US-09-105-390-2
34	37.2	2.3	670	4	US-09-252-991A-10938
35	37.2	2.3	1785	4	US-09-252-991A-11178
36	37.2	2.3	2091	4	US-09-252-991A-11076
37	36.8	2.3	7012	4	US-09-221-017B-173
38	36.4	2.3	1269	4	US-09-252-991A-8698
39	36.4	2.3	1443	4	US-09-252-991A-3554
40	36.4	2.3	1830	4	US-09-252-991A-3563
41	36.4	2.3	1926	4	US-09-252-991A-3531
42	36.4	2.3	1944	4	US-09-252-991A-9081
43	36.4	2.3	3494	3	US-09-139-802-200
44	36.4	2.3	3494	4	US-09-659-786-200
45	36.4	2.3	7218	1	US-08-232-463-14

ALIGNMENTS

RESULT 1

US-08-943-714-1  
; Sequence 1, Application US/08943714  
; Patent No. 6187578

; GENERAL INFORMATION:  
; APPLICANT: Blinkovsky, Alexander

; APPLICANT: Berkla, Randy  
; APPLICANT: Rey, Michael

; APPLICANT: Golightly, Elizabeth  
; APPLICANT: Klotz, Alan

; APPLICANT: Mathisen, Thomas Erik  
; APPLICANT: Dambmann, Claus

; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids  
; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NO. 61875780 No. 6187578disk of No. 6187578th America, Inc.

; STREET: 405 Lexington Avenue

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10174

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/943,714

; FILING DATE: 03-OCT-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J

; REGISTRATION NUMBER: 33,728

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1662 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: Genomic DNA

; FEATURE:

; NAME/KEY: Coding Sequence

; LOCATION: 1..1662

; OTHER INFORMATION:

100%

US-08-943-714-1

Query Match 100.0%; Score 1608; DB 3; Length 1662;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTCCAGGAGTACCGGGGTCGGTCTGGTAGAGACAGTACCCCAAGAACCCACCGGG 60  
DB 55 CTTCCAGGAGTACCGGGGTCGGTCTGGTAGAGACAGTACCCCAAGAACCCACCGGG 114  
QY 61 GTCAGAGCTTTTACACCGCAACAAATGTCACCATCGGTACAGAGAACCCGGGGCAGAG 120  
DB 115 GTCAAGACTCTTCAACCGCAACAAATGTCACCATCGGTACAGAGAACCCGGGGCAGAG 174  
QY 121 GGGGTCTGGCAGACTACCCGGGTGTCAAATCTCTCTGATATGTCGACACTCTCCC 180  
DB 175 GGGGTCTGGCAGACTACCCGGGTGTCAAATCTCTCTGATATGTCGACACTCTCCC 234  
QY 181 GAGTCCCATACCTTCTTCTGGTCTTTCGAAGCCAGACATAACCCAGAAACTGCACCTATC 240  
DB 235 GAGTCCCATACCTTCTTCTGGTCTTTCGAAGCCAGACATAACCCAGAAACTGCACCTATC 294  
QY 241 ACATTTGGTGTGAATGGTGGCCCTGGAGCGATTCCTTGTATCGGTCTCTTCGAAGAGTTG 300  
DB 295 ACATTTGGTGTGAATGGTGGCCCTGGAGCGATTCCTTGTATCGGTCTCTTCGAAGAGTTG 354  
QY 301 GGGCCTTGCCATGTCATTCGACTTTTGTGATGACTACATCAACCTCTACTCTGTTGAACGAG 360  
DB 355 GGGCCTTGCCATGTCATTCGACTTTTGTGATGACTACATCAACCTCTACTCTGTTGAACGAG 414  
QY 361 GTCTCCCAATTTACATTCCTTCCAGCCATTTGGGAGTCGGCTTTTCATATAGATACG 420  
DB 415 GTCTCCCAATTTACATTCCTTCCAGCCATTTGGGAGTCGGCTTTTCATATAGATACG 474  
QY 421 GTTGATGGTCCATTAACCTGTAACTTGGGGTCTCTGAAATTCGAGCTTTCGAGGAGTT 480  
DB 475 GTTGATGGTCCATTAACCTGTAACTTGGGGTCTCTGAAATTCGAGCTTTCGAGGAGTT 534  
QY 481 CAGGGCCGGTACCCCAACCAATGATGCCACTCTGATGACTATACCAATCTTGGCCGACAG 540  
DB 535 CAGGGCCGGTACCCCAACCAATGATGCCACTCTGATGACTATACCAATCTTGGCCGACAG 594  
QY 541 GCGCCTTGGGAGATCCCTGCAAGAGATTCCTTAGTGGACTACCTAGCTGGACTCAGGGTG 600  
DB 595 GCGCCTTGGGAGATCCCTGCAAGAGATTCCTTAGTGGACTACCTAGCTGGACTCAGGGTG 654  
QY 601 CAGTCTAAGGACTTCAGTCTATGACGAGAGCTATGAGGGGACTATGCTCTGCAATC 660  
DB 655 CAGTCTAAGGACTTCAGTCTATGACGAGAGCTATGAGGGGACTATGCTCTGCAATC 714  
QY 661 TTCAATCATTTTACGAGCAGAAATGAGAGATTCGCAACGGTAGTGTATATGGTGTTCAG 720  
DB 715 TTCAATCATTTTACGAGCAGAAATGAGAGATTCGCAACGGTAGTGTATATGGTGTTCAG 774  
QY 721 CTTAATTTCAACTCTCTGGGAAATTAATCAACGGATCATPCAGAGGGGATCCAGGCCCT 780  
DB 775 CTTAATTTCAACTCTCTGGGAAATTAATCAACGGATCATPCAGAGGGGATCCAGGCCCT 834  
QY 781 TACTACCTGATTCGCTGTGAACATACCTTACGGTATCAGGGTGTCAACGAGACGCTC 840  
DB 835 TACTACCTGATTCGCTGTGAACATACCTTACGGTATCAGGGTGTCAACGAGACGCTC 894  
QY 841 TACAACATACATGAAGTTTGCACCAACCAATGCTGTCAGGAGTTTGTATTTCCACC 900  
DB 895 TACAACATACATGAAGTTTGCACCAACCAATGCTGTCAGGAGTTTGTATTTCCACC 954  
QY 901 TGCAAAACAGACAAACCGCACCGCATTAGCTAGCTACGCGCTCTGCGCGGAAAGCCACCAAC 960  
DB 955 TGCAAAACAGACAAACCGCACCGCATTAGCTAGCTACGCGCTCTGCGCGGAAAGCCACCAAC 1014  
QY 961 ATGTGCGGGGACAATGTTGAGGGGCACTACGCTTTGCTGGTGGTGGTGTGTATGAT 1020  
DB 1015 ATGTGCGGGGACAATGTTGAGGGGCACTACGCTTTGCTGGTGGTGGTGTGTATGAT 1074

RESULT 2

US-09-640-305-3  
; Sequence 3, Application US/09640305  
; Patent No. RE37447

GENERAL INFORMATION:

APPLICANT: Fleer, Reinhard  
Fournier, Alain  
Yeh, Patrice

TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR  
PREPARATION AND USE

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Rhone-Poulenc Rorer Inc.

STREET: 500 Arcola Rd. 3C43

CITY: Collegeville

STATE: PA

COUNTRY: USA

ZIP: 19002

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/640,305

FILING DATE: 16-Aug-2000

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/360,673

FILING DATE: 06-FEB-1995

APPLICATION NUMBER: WO PCT/FR93/00623

FILING DATE: 23-JUN-1993

COMPUTER: IBM PC compatible

ADDRESSEE: NO. 35941198 NO. 3594119disk of NO. 5594119th America, Inc.  
STREET: 405 Lexington Avenue. Suite 6400

; COMPUTER: IBM PC compatible

CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/309,341  
FILING DATE: 16-SEP-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lowney, Karen A.  
REGISTRATION NUMBER: 31,274  
REFERENCE/DOCKET NUMBER: 4247,000-US  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2002 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Aspergillus niger  
FEATURE:  
NAME/KEY: intron  
LOCATION: 349..411  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join (348..412)  
US-08--309-341-3

Query Match 5.2%; Score 84; DB 1; Length 2002;  
Best Local Similarity 55.5%; Pred. No. 2,8e-15;  
Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

```

169 GACACCTCTCCCGAGTCCCATACCTCTTCTGTTCTTCGAAAGCCAGACATACCCAGAA 228
167 GACACCTCTCCCGAGTCCCATACCTCTTCTGTTCTTCGAAAGCCAGACATACCCAGAA 228
617 GACCAAGGACAAGACAAAGCATGTGTTCTACTGTGTTCTTCGAGTCTGCAATGACCCGAG 675
167 GACCAAGGACAAGACAAAGCATGTGTTCTACTGTGTTCTTCGAGTCTGCAATGACCCGAG 675
229 ACTGCACCTATCACATGTTGSGTTGAATGTTGGCCCTGGAAAGCGATTCTTTTGATCGGTCTC 288
229 ACTGCACCTATCACATGTTGSGTTGAATGTTGGCCCTGGAAAGCGATTCTTTTGATCGGTCTC 288
677 AATGACCCTTGTCTCTGTGCTGAACGTGGGCCCTGGATGCTCTTCCCTACCGGTCTT 736
677 AATGACCCTTGTCTCTGTGCTGAACGTGGGCCCTGGATGCTCTTCCCTACCGGTCTT 736
289 TTCGAGAGATTGGGCCCTTGGCCATGTCATATTCGACATTTTGATGACATACATCAACCCCTAC 348
289 TTCGAGAGATTGGGCCCTTGGCCATGTCATATTCGACATTTTGATGACATACATCAACCCCTAC 348
737 TTCATGGAGCTGGCCCTAGCAGCATCAACAAGAAATCCAGCCGGGTCTCAACAGCACTAC 796
737 TTCATGGAGCTGGCCCTAGCAGCATCAACAAGAAATCCAGCCGGGTCTCAACAGCACTAC 796
349 TCGTGGAAAGCAGGTCTCCCAATTTACTATTCTCTCCAGCCATTGGGAGTGGCGTTTCA 408
349 TCGTGGAAAGCAGGTCTCCCAATTTACTATTCTCTCCAGCCATTGGGAGTGGCGTTTCA 408
797 GCTTGGAACTCCAAACGCTCCGTTGATCTTCCTTGACCAAGCCTGTCAACGTCGGTTACTCT 856
797 GCTTGGAACTCCAAACGCTCCGTTGATCTTCCTTGACCAAGCCTGTCAACGTCGGTTACTCT 856
409 TATAGTGATACGGTGTGATCGGTGCCAATAACCCCTGTACTGGGGTCTGCGA 458
409 TATAGTGATACGGTGTGATCGGTGCCAATAACCCCTGTACTGGGGTCTGCGA 458
857 TACAGCAACTCTGCTGTACGCGACACCGTGTGCTGGCAAGGACGCTTA 906
857 TACAGCAACTCTGCTGTACGCGACACCGTGTGCTGGCAAGGACGCTTA 906

RESULT 5
US-08-608-267-3
; Sequence 3, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5688663o No. 5688663disk of No. 5688663th America, Inc.

```





```
RESULT 8
US-08-967-149-3
; Sequence 3, Application US/08967149
; Patent No. 5939305
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59393050 No. 5939305disk of No. 5939305th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: 28-FEB-1996
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 202 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 349..411
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (348..412)
;
US-08-967-149-3
Query Match 5.2%; Score 84; DB 2; Length 2002;
Best Local Similarity 55.5%; Pred. No. 2.8e-15;
Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 169 GACACTCTCCGAGTCCCATACCTCTCTCTGTTCTTCGAGAGCCAGACATAACCCAGAA 228
DB 617 GACACGAGAGACGACGAAGCATCTGTCTACTGTCTTCGAGTCTCGCAATGACCCCGAG 676
QY 229 ACTGCACCTATCATATTTGGTTGAATGGTGGCCCTGGAAGCAATCTTTTGATCGGTCTC 288
DB 677 ATGACCCCTGTGTTCTGTGGCTGACGGTGGCCCTGTGATGCTCTTCCCTACCGGCTCT 736
QY 289 TTCGAAGAGTTGGGCGCTTGCATGTCATTTGACATTTTGTGATGACTACATCAACCTCAC 348
DB 737 TTATGGAGCTCGGCCCTAGCAGCATCAACAAGAAGATCCAGCCGGCTCTACACGACTAC 796
QY 349 TCGTGGAGAGGATCTCCCAATTACTATTCTCTGTCCTCCAGCCATTGGGATCGCGCTTTCA 408

Db 797 GCTTGAACCTCCACGCGTCCCGTGATCTTCTTGACCAAGCTGTCAAGTTCGGTACTCT 856
QY 409 TATAGTATACGGTTGATGGGTCCCATTAACCCCTGTAACTGGGGTCTGTGA 458
Db 857 TACAGCAACTCTGCTGTCTCAGCAGACACCGTGTGCTGCGCAGGACGCTCA 906

RESULT 9
US-08-309-341-1
; Sequence 1, Application US/08309341
; Patent No. 5594119
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5594119o No. 5594119disk of No. 5594119th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,341
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 572..632
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (571..633)
;
US-08-309-341-1
Query Match 4.7%; Score 76; DB 1; Length 2068;
Best Local Similarity 55.8%; Pred. No. 7.5e-13;
Matches 145; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 199 TGGTCTTCGAGCCAGACATAACCCAGAACTGCACCATTCACATTTGTTGATGGT 258
DB 633 TGGTCTTCGAGTCTGCAATGACCCCGAGATGATCCCGTTGTTCTGTGGTGAACGGT 692
QY 259 GGCCTCGGAAGCGATTTTGTGATCGGTCTCTTGAAGAGTTGGGCCCTTGCCATGCAAT 318
DB 693 GGCCTCGGTTGCTCTTCCCTCACCCTCTCTCATGAGACTTGGCCCTAGCAGCATCAAC 752
QY 319 TCGACTTTTGTGACTATCATCAACCCCTCACTCGTGGAGAGGAGTCTTCCAAATTTACTTC 378
DB 753 AAGAAGATCCAGCCGGTCTACATGACTAGCTTGAAGTCTACGCTTGAAGTCCCAACCGCTCGGTGATCTTC 812
```

199	TGTTCTTCGAAGCCAGACATACCCAGAAACTGCACCTATCATCAATTCIGGTTGAATGGT	258
633	TGTTCTTCAGATCTCGCAATCAACCCGAGAAATGATCCCGTTGTCTTGCTTGCTGGCTGAACGGT	692
259	GGCCCTGGAAAGCGATCTTTTGATCGGTCTCTTCGAAGAGTTGGGCCCTTGGCATTGTCAAT	318
693	GGCCCTGGGTGTCCTTCCTCAACCGGTCTCTTCATGTGAGCTTGGCCCTAGCAGCAATCAAC	752





```
;
;
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 8648.44US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/332-5300
; TELEFAX: 612/332-9081
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2632 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 696...2291
; OTHER INFORMATION:
;
US-08-329-892B-32

Query Match          4.1%; Score 66.2; DB 3; Length 2632;
Best Local Similarity 53.2%; Pred. No. 7.7e-10;
Matches 140; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 187 CATACCTCTCTCTGGTTCITCGAAGCCGACACATAACCCAGAACTGCACCTATCACATTG 246
   ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1113 CATTCTCTCTTTGGACTTTTGAAGTAGAAGGATCCTGCAAGGATCGGTCATCCTT 1172

QY 247 TGGTTGAATGGTGGCCCTGGAAGGATTCTTGTGATCGGTCTCTTCGAAGAGTTGGGCCCT 306
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1173 TGGTTGAACGGGGTCCAGGTTGTTCTTCACIACCGGGCTGTTCTTGTATTAGGACCC 1232

QY 307 TGCCATGTCATTCGACTTTTGTAGTACTACATCAACCCICACCTGCGTGGACAGGTCCTCC 366
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1233 TCATCCATTGGACCTGATTGAAACCCATCGGAACCTTACTCTTGGACAGCAATGCC 1292

QY 367 AATTACTATTCCTGTCGCCAGCCATTGGGAGTCGGCTTTTCATATAGTATACGGTTGAT 426
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1293 ACCGTGATCTTCCTTGACCAACCCGTCAACGTTGGGTCTCGTATCCGGGTCTCAGGT 1352

QY 427 GGGTCCATTAAACCTGTAACTGG 449
   | | | | | | | | | | | | | |
Db 1353 GTTCCACACTGTCGCCGCTGG 1375
```

Search completed: September 16, 2003, 14:45:40  
Job time : 150 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 14:45:47 ; Search time 1475 Seconds  
(without alignments)  
2681.728 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662

Perfect score: 1608

Sequence: 1 cttccaggaaagtacacggc.....gcattgtccagttgtgtatg 1608

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1660708 seqs, 122959015 residues

Total number of hits satisfying chosen parameters: 3321416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70.6	4.4	2218	14	US-10-206-619-1
2	56.4	3.5	1814	10	US-09-736-457-319
3	56.4	3.5	1814	10	US-09-902-941-319
4	56.4	3.5	1814	10	US-09-849-626-319
5	56.4	3.5	1814	11	US-09-476-300-319
6	56.4	3.5	1814	12	US-10-113-872-319
7	56.4	3.5	1814	14	US-10-017-754-319
8	56.4	3.5	1815	10	US-09-880-107-2296
9	54.8	3.4	2864	9	US-09-925-302-331
10	47.6	3.0	390	9	US-09-790-389-7
11	47.6	3.0	1551	13	US-10-084-018-4
12	47.6	3.0	1684	9	US-09-729-674-109
13	47.6	3.0	1697	11	US-09-796-753-39
14	47.6	3.0	2076	10	US-09-909-320-163
15	47.6	3.0	2076	10	US-09-909-088B-163
16	47.6	3.0	2076	10	US-09-905-291A-163

```

17 47.6 3.0 2076 10 US-09-902-853-163 Sequence 163, App
18 47.6 3.0 2076 10 US-09-907-824-163 Sequence 163, App
19 47.6 3.0 2076 10 US-09-907-841-163 Sequence 163, App
20 47.6 3.0 2076 11 US-09-904-011-163 Sequence 163, App
21 47.6 3.0 2076 11 US-09-906-742-163 Sequence 163, App
22 47.6 3.0 2076 11 US-09-906-838-163 Sequence 163, App
23 47.6 3.0 2076 11 US-09-907-613-163 Sequence 163, App
24 47.6 3.0 2076 11 US-09-907-942-163 Sequence 163, App
25 47.6 3.0 2076 11 US-09-904-859-163 Sequence 163, App
26 47.6 3.0 2076 11 US-09-909-204-163 Sequence 163, App
27 47.6 3.0 2076 11 US-09-904-820-163 Sequence 163, App
28 47.6 3.0 2076 11 US-09-904-786-163 Sequence 163, App
29 47.6 3.0 2076 11 US-09-906-646-163 Sequence 163, App
30 47.6 3.0 2076 11 US-09-906-700-163 Sequence 163, App
31 47.6 3.0 2076 11 US-09-903-786-163 Sequence 163, App
32 47.6 3.0 2076 11 US-09-902-903-163 Sequence 163, App
33 47.6 3.0 2076 11 US-09-903-749A-163 Sequence 163, App
34 47.6 3.0 2076 11 US-09-904-119-163 Sequence 163, App
35 47.6 3.0 2076 11 US-09-904-956-163 Sequence 163, App
36 47.6 3.0 2076 11 US-09-902-736-163 Sequence 163, App
37 47.6 3.0 2076 11 US-09-907-794-163 Sequence 163, App
38 47.6 3.0 2076 11 US-09-903-943-163 Sequence 163, App
39 47.6 3.0 2076 11 US-09-904-462-163 Sequence 163, App
40 47.6 3.0 2076 11 US-09-907-925-163 Sequence 163, App
41 47.6 3.0 2076 11 US-09-902-692-163 Sequence 163, App
42 47.6 3.0 2076 11 US-09-903-520-163 Sequence 163, App
43 47.6 3.0 2076 11 US-09-903-056-163 Sequence 163, App
44 47.6 3.0 2076 11 US-09-909-064-163 Sequence 163, App
45 47.6 3.0 2076 11 US-09-904-553-163 Sequence 163, App

```

#### ALIGNMENTS

#### RESULT 1

```

US-10-206-619-1
; Sequence 1, Application US/10206619
; Publication No. US20030108908A1
; GENERAL INFORMATION:
; APPLICANT: Korea Research Institute of Bioscience and Biotechnology
; APPLICANT: Dong Kook Pharmaceutical Co.
; APPLICANT: Rhee, Sangki
; APPLICANT: Choi, Eulsung
; APPLICANT: Kang, Hyunah
; APPLICANT: Sohn, Junghoon
; APPLICANT: Bae, Junghoon
; APPLICANT: Kim, Moowoong
; APPLICANT: Agaphonov, Michasel
; TITLE OF INVENTION: Hansenula polymorpha mutants and process for the preparation o
; TITLE OF INVENTION: recombinant proteins using the same
; FILE REFERENCE: 4220-116 US
; CURRENT APPLICATION NUMBER: US/10/206,619
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US/09/674,617
; PRIOR FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: PCT/KR00/00173
; PRIOR FILING DATE: 2000-03-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2218
; TYPE: DNA
; ORGANISM: Hansenula polymorpha DLI
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: 067174
; DATABASE ENTRY DATE: 1996-08-17
; RELEVANT RESIDUES: (1)..(2218)
US-10-206-619-1

```

Query Match 4.4%; Score 70.6; DB 14; Length 2218;  
Best Local Similarity 53.4%; Pred. No. 2.8e-12;  
Matches 148; Conservative 0; Mismatches 129; Indels 0; Gaps 0;





; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 319

; LENGTH: 1814

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-849-626-319

Query Match 3.5%; Score 56.4; DB 10; Length 1814;  
Best Local Similarity 52.5%; Pred. No. 1.5e-07;  
Matches 148; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

```
QY 139 CCGGGTGTCAAATCTTACTCTGGATATGTCGACACCTCTCCGAGTCCCATACCTTCTTC 198
Db 139 CCGTCTTTCGGCCAGTACTCCGGTACCTCAAAAGCTCCGGCTCCAGACCTCCCACTAC 198
QY 199 TGGTCTTTCGAAGCCAGACATAACCCAGAACTCCACCTATCACTATGTTGGTGAATGGT 259
Db 199 TGGTCTTTCGAAGCCAGACATAACCCAGAACTCCACCTATCACTATGTTGGTGAATGGT 259
QY 259 GGCCCTGGAAGCGATTCTTTGATCGGTCTCTTGAAGAGTGGGCCCTTCCCATGTCAAT 318
Db 259 GGTCCGGCTGCGAGTCACTAGATGGCTCTTACAGAGCATGGCCCTTCTTGGTCCAG 318
QY 319 TCGACTTTTGTGACT---ACATCAACCTTCACTCGTGAACGAGGTCTCCAAATTTACTA 375
Db 319 CCAGATGGTGTACCCCTGAGTACAAACCCCTATCTTGGAACTCTGATGGCAATGTGTTA 378
QY 376 TTCTCTGTCCAGCATTTGGAGTCGGCTTTTCATATAGTAT 417
Db 379 TACTTGGAGTCCCACTGGGTGGGCTTCTCTACTCCGAT 420
```

## RESULT 5

US-09-476-300-319

; Sequence 319, Application US/09476300

; Publication No. US20030125245A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Bangur, Chaitanya S.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.478C3

; CURRENT APPLICATION NUMBER: US/09/476,300

; CURRENT FILING DATE: 1999-12-30

; NUMBER OF SEQ ID NOS: 785

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 319

; LENGTH: 1814

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-476-300-319

Query Match 3.5%; Score 56.4; DB 11; Length 1814;  
Best Local Similarity 52.5%; Pred. No. 1.5e-07;  
Matches 148; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

```
QY 139 CCGGGTGTCAAATCTTACTCTGGATATGTCGACACCTCTCCGAGTCCCATACCTTCTTC 198
Db 139 CCGTCTTTCGGCCAGTACTCCGGTACCTCAAAAGCTCCGGTCCAGACCTCCCACTAC 198
QY 199 TGGTCTTTCGAAGCCAGACATAACCCAGAACTCCACCTATCACTATGTTGGTGAATGGT 258
Db 199 TGGTCTTTCGAAGCCAGACATAACCCAGAACTCCACCTATCACTATGTTGGTGAATGGT 258
QY 259 GGCCCTGGAAGCGATTCTTTGATCGGTCTCTTGAAGAGTGGGCCCTTCCCATGTCAAT 318
Db 259 GGTCCGGCTGCGAGTCACTAGATGGCTCTTACAGAGCATGGCCCTTCTTGGTCCAG 318
QY 319 TCGACTTTTGTGACT---ACATCAACCTTCACTCGTGAACGAGGTCTCCAAATTTACTA 375
Db 319 CCAGATGGTGTACCCCTGAGTACAAACCCCTATCTTGGAACTCTGATGGCAATGTGTTA 378
QY 376 TTCTCTGTCCAGCATTTGGAGTCCGGCTTTTCATATAGTAT 417
```

```
Db 379 TACCTGGAGTCCCACTGGGTGGGTTCTCCTACTCCGAT 420
```

## RESULT 6

US-10-113-872-319

; Sequence 319, Application US/10113872

; Publication No. US20030170255A1

; GENERAL INFORMATION:

; APPLICANT: Watanabe, Yoshihiro

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Sleath, Paul R.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Carter, Darrick

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.478C19

; CURRENT APPLICATION NUMBER: US/10/113,872

; CURRENT FILING DATE: 2002-03-28

; NUMBER OF SEQ ID NOS: 2011

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 319

; LENGTH: 1814

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-113-872-319

Query Match 3.5%; Score 56.4; DB 12; Length 1814;  
Best Local Similarity 52.5%; Pred. No. 1.5e-07;  
Matches 148; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

```
QY 139 CCGGGTGTCAAATCTTACTCTGGATATGTCGACACCTCTCCGAGTCCCATACCTTCTTC 198
Db 139 CCGTCTTTCGGCCAGTACTCCGGTACCTCAAAAGCTCCGGTCCAGACCTCCCACTAC 198
QY 199 TGGTCTTTCGAAGCCAGACATAACCCAGAACTCCACCTATCACTATGTTGGTGAATGGT 258
Db 199 TGGTCTTTCGAAGCCAGACATAACCCAGAACTCCAGAGGATCCGAGACAGCCCTTGTGCTTGGCTCAATGGG 258
QY 259 GGCCCTGGAAGCGATTCTTTGATCGGTCTCTTGAAGAGTGGGCCCTTCCCATGTCAAT 318
Db 259 GGTCCGGCTGCGAGTCACTAGATGGGCTCTTACAGAGCATGGCCCTTCTTGGTCCAG 318
QY 319 TCGACTTTTGTGACT---ACATCAACCTTCACTCGTGAACGAGGTCTCCAAATTTACTA 375
Db 319 CCAGATGGTGTCAACCTCCGAGTACAAACCCCTATCTTGGAACTCTGATGGCAATGTGTTA 378
QY 376 TTCTCTGTCCAGCATTTGGAGTCCGGCTTTTCATATAGTAT 417
Db 379 TACCTGGAGTCCCACTGGGTGGGCTTCTCCTACTCCGAT 420
```

## RESULT 7

US-10-017-754-319

; Sequence 319, Application US/10017754

; Publication No. US20030054363A1

; GENERAL INFORMATION:

; APPLICANT: Henderson, Robert A.

; APPLICANT: Wang, Tongtong

; APPLICANT: Watanabe, Yoshihiro

; APPLICANT: Johnson, Jeffrey C.

; APPLICANT: Retter, Marc W.

; APPLICANT: Marnerakis, Margarita

; APPLICANT: Carter, Darrick

; APPLICANT: Fanger, Gary R.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: McNabb, Andria

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.478C18

```
; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 319
; LENGTH: 1814
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-017-754-319

Query Match          3.5%; Score 56.4; DB 14; Length 1814;
Best Local Similarity 52.5%; Pred. No. 1.5e-07;
Matches 148; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

QY 139 CGGGGTGTAATCTTACTCTGATATGTCGACACCTCTCCCGAGTCCCATACCTTCTTC 198
Db 139 CGGTCTTCCGCGAGTACTCCGCTACCTCAAAAGCTCCGGTCCAAAGCACCCTCCACTAC 198
QY 199 TGGTTCTCGAAGCCAGACATACCCAGAACTGCACCTATACATTTGGTGGTGAATGGT 258
Db 199 TGGTTCTCGAGTCCAGAGGATCCCGAGACAGCCCTGTGTGCTTTGGCTCAATGGG 258
QY 259 GGCCTCGAAGCGAATCTTTTGATCGGTCTCTTCGAAGAGTTGGGCCCTTGCCTATGCAAT 318
Db 259 GGTCCGGCTGCGAGTCACTAGATGGCTCTCTCACAGAGCATGGCCCTCTCTGGTCCAG 318
QY 319 TCGACTTTTGATGACT--ACATCAACCTCCTGCTGGAACGAGTCTCCAAATTTACTA 375
Db 319 CCAGATGGTGTCAACCTCGGAGTACAAACCCCTATCTTGGAACTGATGCAATGCTTA 378
QY 376 TTCTCTCCAGCAGCAITGGGAGTCGGCTTTTCATATAGTAT 417
Db 379 TACCTGGAGTCCCGAGCTGGGGTGGCTTCTCTACTCCGAT 420

RESULT 9
US-09-925-302-331
; Sequence 331, Application US/09925302
; Patent No. US2002044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 331
; LENGTH: 2864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2850)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2858)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2860)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-331

Query Match          3.4%; Score 54.8; DB 9; Length 2864;
Best Local Similarity 52.1%; Pred. No. 6.6e-07;
Matches 147; Conservative 0; Mismatches 132; Indels 3; Gaps 1;

QY 139 CCGGGTGTCAAATCCTACTCTGGATATGTGACACCTCTCCCGAGTCCCATACCTTCTTC 198
Db 171 CCGTCTTTCGCGAGTACTCCGGCTACTCCTCAAAGGCTCCGGCTCCAGCACCCTCCACTAC 230
QY 199 TGGTCTCTCGAAGCCAGACATAACCCAGAACTGCACCTATCACATTTGGTGTGAATGGT 258
Db 231 TGGTTTGGAGTCCCGAGAGATCCCGAGACAGCCCTGTGGTCTTGGCTCAATGGG 290
QY 259 GGCCTCGAAGCGAATTTTGTATGTCGTCTCTTCGAAGAGTTGGGCCCTTGGCATGTCAAT 318
Db 291 GGTCCCGCTGCGAGTCACTAGATGGGTCTCTCACAGAGCATGGCCCTCTCTGGTCCAG 350
QY 319 TCGACTTTTGATGACT--ACATCAACCTCCTGCTGGAACGAGTCTCCAAATTTACTA 375
Db 351 CCAGATGGTGTCAACCTCGGAGTACAAACCCCTATCTTGGAACTGATTTGCCAATGTGTTA 410
QY 376 TTCTCTCCAGCAGCAITGGGAGTCGGCTTTTCATATAGTAT 417
Db 411 TACCTGGAGTCCCGAGCTGGGGTGGCTTCTCTACTCCGAT 452

RESULT 10
US-09-790-399-7

; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 319
; LENGTH: 1814
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-017-754-319

Query Match          3.5%; Score 56.4; DB 14; Length 1815;
Best Local Similarity 52.5%; Pred. No. 1.5e-07;
Matches 148; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

QY 139 CCGGGTGTCAAATCCTACTCTGATATGTCGACACCTCTCCCGAGTCCCATACCTTCTTC 198
Db 139 CGGTCTTCCGCGAGTACTCCGCTACCTCAAAAGCTCCGGTCCAAAGCACCCTCCACTAC 198
QY 199 TGGTTCTCGAAGCCAGACATACCCAGAACTGCACCTATACATTTGGTGGTGAATGGT 258
Db 199 TGGTTCTCGAGTCCAGAGGATCCCGAGACAGCCCTGTGTGCTTTGGCTCAATGGG 258
QY 259 GGCCTCGAAGCGAATCTTTTGATCGGTCTCTTCGAAGAGTTGGGCCCTTGCCTATGCAAT 318
Db 259 GGTCCGGCTGCGAGTCACTAGATGGCTCTCTCACAGAGCATGGCCCTCTCTGGTCCAG 318
QY 319 TCGACTTTTGATGACT--ACATCAACCTCCTGCTGGAACGAGTCTCCAAATTTACTA 375
Db 319 CCAGATGGTGTCAACCTCGGAGTACAAACCCCTATCTTGGAACTGATGCAATGCTTA 378
QY 376 TTCTCTCCAGCAGCAITGGGAGTCGGCTTTTCATATAGTAT 417
Db 379 TACCTGGAGTCCCGAGCTGGGGTGGCTTCTCTACTCCGAT 420

RESULT 8
US-09-980-107-2296
; Sequence 2296, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2296
; LENGTH: 1815
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M22960
US-09-980-107-2296

Query Match          3.5%; Score 56.4; DB 10; Length 1815;
Best Local Similarity 52.5%; Pred. No. 1.5e-07;
Matches 148; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

QY 139 CCGGGTGTCAAATCCTACTCTGATATGTCGACACCTCTCCCGAGTCCCATACCTTCTTC 198
Db 139 CGGTCTTCCGCGAGTACTCCGCTACCTCAAAAGCTCCGGTCCAAAGCACCCTCCACTAC 198
QY 199 TGGTTCTCGAAGCCAGACATAACCCAGAACTGCACCTATACATTTGGTGGTGAATGGT 258
Db 199 TGGTTCTCGAGTCCAGAGGATCCCGAGACAGCCCTGTGTGCTTTGGCTCAATGGG 258
QY 259 GGCCTCGAAGCGAATCTTTTGATCGGTCTCTTCGAAGAGTTGGGCCCTTGCCTATGCAAT 318
Db 259 GGTCCGGCTGCGAGTCACTAGATGGCTCTCTCACAGAGCATGGCCCTCTCTGGTCCAG 318
QY 319 TCGACTTTTGATGACT--ACATCAACCTCCTGCTGGAACGAGTCTCCAAATTTACTA 375
Db 319 CCAGATGGTGTCAACCTCGGAGTACAAACCCCTATCTTGGAACTGATGCAATGCTTA 378
QY 376 TTCTCTCCAGCAGCAITGGGAGTCGGCTTTTCATATAGTAT 417
Db 379 TACCTGGAGTCCCGAGCTGGGGTGGCTTCTCTACTCCGAT 420
```

RESULT 11  
US-10-084-018--4  
Sequence 4, Application US/10084018  
Publication No. US20020160499A1  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
Hawkins, Phillip R.  
Hillman, Jennifer L.  
Lal, Preeti

Query Match	3.0%	Score	47.6	DB	13	Length	1551
Best Local Similarity	46.9%	Prod. No.	0.00012				
Matches	149	Conservative	0	Mismatches	169	Indels	0
Gaps	0						
166	QY	gTCGACACCTCTCCCGAGTCCCATACCTCTCTCTGTTGTTCTTCGAGCCGACATAAACCA	225				
320	Ddb	GTGATATAGACTTACAAACAGCAACCTCTCTCTGTTGTTCTTCCAGCTCAGATACAGCCA	379				
226	QY	GAACCTGCACCTATACATATGTTGTTGAATGGTGGCCCTGGAAGCGATTCTTTTCGCGGT	285				
380	Ddb	GAAGAIGCCCCAGTAGTTCCTGCGCTACAGGGTGGCCGGAGGTTCATCCATGTTWGGG	439				
286	QY	CTCTTCGAAGAGTTGGGCCCTTGGCCATGTCAAATTCGACTTTTATGACTACATCAACCT	345				
440	Ddb	CTCTTTGGGAACATGGGCGCTATGTTGTCAAGTAACATGACCTTGGCGTACAGAGAC	499				
346	QY	CACCTGGTGAAGAGGTCTCCAAATTACTATTTCCTGTCGCCAGCCATGGGAGTCGCTTT	405				
500	Ddb	TTCCTCTGGACACAAGSTCTCCCATGCTTTCATATGACAATCCAGTGGCGACAGCTTC	559				
406	QY	TCATATAGTATPACGGTTGATGGGTCCATTAACTTAACTGGGTCGCTCGAAAATCG	465				
560	Ddb	AGTTTTACTGATATACCCACGGATATGCAGTCAATGAGGACGATGTACACGGGATTTA	619				
466	QY	AGCTTTGCAGGAGTTTCAG	483				
620	Ddb	TACAGTGCATTAATTCAG	637				

RESULT 12

```
US-09-729-674-109
; Sequence 109, Application US/09729674
; Patent No. US2001003335A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Meirberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6055-64X
; CURRENT APPLICATION NUMBER: US/09/729,674
; CURRENT FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: 09/539,330
; PRIOR FILING DATE: 2000-03-30
; NUMBER OF SEQ ID NOS: 283
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 109
; LENGTH: 1684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-729-674-109

Query Match      3.0%; Score 47.6; DB 9; Length 1684;
Best Local Similarity 46.9%; Pred. No. 0.00013;
Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 166 GTCGACACCTCTCCGAGTCCCATACCTCTCTCTGTTCTTCGAGGACGACATACACCA 225
Db 332 GTGATAGACTTACACAGCAACCTCTCTCTGTTCTTCGAGGACGACATACACCA 391
QY 226 GAAACTGACCTATFACATGTTGTTGAATGGTGGCCCTGGAACGATCTTTGATCGGT 285
Db 392 GAAGATGCCCGAGTAGTCTCTGGCTACAGGTGGCGGAGGTTCAATCATGTTTGA 451
QY 286 CTCCTGCAAGATGGGCCCTTGCCATGCTCAATTCGACTTTTGATGACTACATCAACCT 345
Db 452 CTCCTTGTGGCAATGGGCCCTTATGTTGTCACAGTACATGACCTTCGTCAGACGAC 511
QY 346 CACCTGTTGGAACGAGTCTCCAAATTTACTATTCCTGTCGCCAGCCATGGGAGTGGCTTT 405
Db 512 TTCCCTGGACCAACAGCTCTCCATGCTTTACATTTGACATCCAGTGGGACAGGCTTC 571
QY 406 TCATATAGTACAGTTGATGGTGGTTCATTAACCTGTAACTGGGTGCTCGAAATTCG 465
Db 572 AGTTTACTGATGATACCCAGGATATGACGTCAATGAGGAGGATGTAGACGGGANTTA 631
QY 466 AGCTTTGACGAGGATTCAG 483
Db 632 TACATGCACTAATTCAG 649

RESULT 13
US-09-796-753-39
; Sequence 39, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30

Query Match      3.0%; Score 47.6; DB 11; Length 1697;
Best Local Similarity 46.9%; Pred. No. 0.00013;
Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 166 GTCGACACCTCTCCGAGTCCCATACCTCTCTCTGTTCTTCGAGGACGACATACCA 225
Db 338 GTGATAGACTTACACAGCAACCTCTCTCTGTTCTTCGAGGACGACATACCA 397
QY 226 GAAACTGACCTATFACATGTTGTTGAATGGTGGCCCTGGAACGATCTTTGATCGGT 285
Db 398 GAAGATGCCCGAGTAGTCTCTGGCTACAGGTGGCGGAGGTTCAATCATGTTTGA 457
QY 286 CTCCTGCAAGATGGGCCCTTGCCATGCTCAATTCGACTTTTGATGACTACATCAACCT 345
```

Db 458 CTCCTTGTGGAACATGGGCGCTTAAGTGTGTCAAGTAACATGACCTGTGGTGACAGAC 517  
 QY 346 CACTCGTGTGAAGAGGTCTCCAAATTACTATTCTCTCCAGCCATGGGAGTCGGCTTT 405  
 Db 518 TTCCCTGTGAACACACACAGCGTCTCCAGCGCTTTACATTTGACAATCCAGTGGGACAGCGTTC 577  
 QY 406 TCATATAGTATGATACAGTGTGATGGGTGTCATTAACCCCTGTAACCTGGGTGTCGAAAATTCG 465  
 Db 578 AGTTTACTGTAGATACCCAGGATATGACGATGAGGAGATGATGACACGGGATTTA 637  
 QY 466 AGCTTTGAGGAGTTCAG 483  
 Db 638 TACAGTGCACATAATTCAG 655

## RESULT 14

US-09-909-320-163  
 ; Sequence 163, Application US/09909320  
 ; Patent No. US20020132240A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Genentech, Inc.  
 ; APPLICANT: Ashkenazi, Avi  
 ; APPLICANT: Botstein, David  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Eaton, Dan L.  
 ; APPLICANT: Ferrara, Napoleone  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Fong, Sherman  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerber, Hanspeter  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, A.  
 ; APPLICANT: Grimaldi, Christopher J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Hillan, Kenneth, J.  
 ; APPLICANT: Kljavin, Ivar J.  
 ; APPLICANT: Mather, Jennie P.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Paoni, Nicholas F.  
 ; APPLICANT: Roy, Margaret Ann  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Wood, William, I.  
 ; APPLICANT: Williams, P. Mickey  
 ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
 ; FILE REFERENCE: 10466-14  
 ; CURRENT APPLICATION NUMBER: US/09/909,320  
 ; CURRENT FILING DATE: 2002-01-04  
 ; PRIOR APPLICATION NUMBER: PCT/US00/04414  
 ; PRIOR FILING DATE: 2000-02-22  
 ; PRIOR APPLICATION NUMBER: US 60/143,048  
 ; PRIOR FILING DATE: 1999-07-07  
 ; PRIOR APPLICATION NUMBER: US 60/145,698  
 ; PRIOR FILING DATE: 1999-07-26  
 ; PRIOR APPLICATION NUMBER: US 60/146,222  
 ; PRIOR FILING DATE: 1999-07-28  
 ; PRIOR APPLICATION NUMBER: PCT/US99/20594  
 ; PRIOR FILING DATE: 1999-09-08  
 ; PRIOR APPLICATION NUMBER: PCT/US99/20944  
 ; PRIOR FILING DATE: 1999-09-13  
 ; PRIOR APPLICATION NUMBER: PCT/US99/21090  
 ; PRIOR FILING DATE: 1999-09-15  
 ; PRIOR APPLICATION NUMBER: PCT/US99/21547  
 ; PRIOR FILING DATE: 1999-09-15  
 ; PRIOR APPLICATION NUMBER: PCT/US99/23089  
 ; PRIOR FILING DATE: 1999-10-05  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28214  
 ; PRIOR FILING DATE: 1999-11-29  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28313  
 ; PRIOR FILING DATE: 1999-11-30

; PRIOR APPLICATION NUMBER: PCT/US99/28564  
 ; PRIOR FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28565  
 ; PRIOR FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: PCT/US99/30095  
 ; PRIOR FILING DATE: 1999-12-16  
 ; PRIOR APPLICATION NUMBER: PCT/US99/30911  
 ; PRIOR FILING DATE: 1999-12-20  
 ; PRIOR APPLICATION NUMBER: PCT/US99/30999  
 ; PRIOR FILING DATE: 1999-12-20  
 ; PRIOR APPLICATION NUMBER: PCT/US00/00219  
 ; PRIOR FILING DATE: 2000-01-05  
 ; NUMBER OF SEQ ID NOS: 423  
 ; SEQ ID NO 163  
 ; LENGTH: 2076  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-909-320-163  
 Query Match 3.0%; Score 47.6; DB 10; Length 2076;  
 Best Local Similarity 46.9%; Pred. No. 0.00014;  
 Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;  
 QY 166 GTCGACACCTCTCCCGAGTCCCATACCTTCTTCTGCTTCTTCCGAGCCAGACATACCCCA 225  
 Db 334 GTGATAAGACHTACACAGCAACCTCTTCTTCTGCTTCTTCCGAGCCAGACATACCCCA 393  
 QY 226 GAACTGACACCTATCAGATTTGGTGAATGGTGGCCCTGGAGCGATCTTTGATCGGT 285  
 Db 394 GAAGATGCCCCAGTAGTTCTCTGGCTACAGGGTGGCCGGAGGTTTCATCCATGTTGA 453  
 QY 286 CTCCTCGAAGAGTGGGCCCTTCCCAITGCAATGCAATTCGATGACTACATCAACCT 345  
 Db 454 CTCCTTGTGGAACATGGGCGCTTATGTTGTCACAAGTAACATGACCTTGGTGACAGAC 513  
 QY 346 CACTCGTGAACGAGGTCTCCAAATTACTATTCTCTCCAGCCATGGGAGTCGGCTTT 405  
 Db 514 TTCCCTGTGACACACACAGCGTCTCCATGCTTACATTTGACAATCCAGTGGGACAGCGTTC 573  
 QY 406 TCATATAGTATGATACGTTGATGGGTGTCATTAACCCCTGTAACCTGGGTGTCGAAAATTCG 465  
 Db 574 AGTTTACTGTAGTATACCCAGGATATGACGATGAGGAGATGATGACACGGGATTTA 633  
 QY 466 AGCTTTGAGGAGTTCAG 483  
 Db 634 TACAGTGCACATAATTCAG 651

## RESULT 15

US-09-909-088B-163  
 ; Sequence 163, Application US/09909088B  
 ; Patent No. US20020146709A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Genentech, Inc.  
 ; APPLICANT: Ashkenazi, Avi  
 ; APPLICANT: Botstein, David  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Eaton, Dan L.  
 ; APPLICANT: Ferrara, Napoleone  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Fong, Sherman  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerber, Hanspeter  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, A.  
 ; APPLICANT: Grimaldi, Christopher J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Hillan, Kenneth, J.  
 ; APPLICANT: Kljavin, Ivar J.  
 ; APPLICANT: Mather, Jennie P.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: 10466-14  
CURRENT APPLICATION NUMBER: US/09/909,088B  
CURRENT FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: PCT/US99/28214  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30  
PRIOR APPLICATION NUMBER: PCT/US99/28564  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/28565  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30095  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: PCT/US99/30911  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US99/30999  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US00/00219  
PRIOR FILING DATE: 2000-01-05  
NUMBER OF SEQ ID NOS: 423  
SEQ ID NO 163  
LENGTH: 2076  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-909-088B-163

Query Match 3.0%; Score 47.6; DB 10; Length 2076;  
Best Local Similarity 46.9%; Pred. No. 0.00014;  
Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 166 GTCGACACCTCTCCGAGTCCCATACCTCTCTCTGCTTCTCGAAGCCAGACATAACCCA 225  
DB 334 GTGAATAAGACCTTACACACACACCTCTCTCTGCTTCTCTCCACCTCAGATACACCCA 393

QY 226 GAACTGCGACCTATACATGCTGATGATGGTGGCCCTGGAAGCCGATCTCTTTGATCGGT 285  
DB 394 GAAGATGCCCCAGTACTCTCTGCTACAGGTGGGCCGCGAGGTTCATCCATGTTTGA 453

QY 286 CTCTTCGAGAGTGGGCCCTTCCCATGCAATTCGACTTTGATGACTACATCAACCCCT 345  
DB 454 CTCCTTGTGAACATGGCCCTTATGTTGTCACAAAGTAACAGACCTTGGTGGACAGAC 513

QY 346 CACTCGTGGAGAGGTCTCCAACTTACTATTCCTGTCAGCCATTTGGAGTGGCTTT 405  
DB 514 TTCCCTGGACCAACAGCTCTCCATGCTTTACATTTGACAATCCAGTGGGCACAGGCTC 573

QY 406 TCATATAGTATCGGTGATGGTCCATTAACCCCTGTAACCTGTAACCTGCTCGTCAAAATCG 465

DB 574 AGTTTACTGATGATACCCACGGATATGCGATGCAATGAGGACGATGACGCGGATTTA 633  
QY 466 AGCTTTGCGAGGAGTTTCAG 483  
DB 634 TACAGTGCCTAATTCAG 651

Search completed: September 16, 2003, 18:38:56  
Job time : 1478 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 14:42:53 ; Search time 4185 Seconds  
(without alignments)  
9338.494 Million cell updates/sec

Title: US-09-712-338-1\_COPY\_55\_1662

Perfect score: 1608

Sequence: 1 cttcaggaagtacacggc.....goatgtccagtgtgtgtatg 1508

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	138.4	8.6	588	9	AW672518
2	116	7.2	721	14	CA747625
3	92.6	5.8	498	12	BM868983
4	86.6	5.4	963	29	CNS0625E

5	86.6	5.4	1049	29	CNS06XWN
6	83.6	5.2	696	13	B0634841
7	80.6	5.0	485	12	BI200653
8	77.6	4.8	445	12	BI187544
9	77.6	4.8	475	9	AW677015
10	77	4.8	1054	29	CNS06NN7
11	74	4.6	729	13	BQ110016
12	72.8	4.5	1002	29	CNS06ZOK
13	72.4	4.5	662	14	CD044859
14	72.2	4.5	682	14	CD054523
15	70.2	4.4	907	14	CD456151
16	69.2	4.3	675	14	CB937789
17	68.2	4.2	1109	14	CD508724
18	68	4.2	580	13	BW197413
19	67.6	4.2	700	12	BI749802
20	67.4	4.2	685	13	BW194502
21	66.2	4.1	457	14	CD039988
22	66.2	4.1	544	14	CD042968
23	66.2	4.1	712	14	CD045123
24	66	4.1	648	10	BE777028
25	66	4.1	653	14	CD258740
26	65.8	4.1	482	9	AV985320
27	65.8	4.1	550	9	AV957236
28	65.8	4.1	566	13	BW276795
29	65.8	4.1	686	13	BW260398
30	65.4	4.1	287	14	W06667
31	64.8	4.0	590	13	BW198196
32	64.2	4.0	551	9	AV976226
33	64.2	4.0	555	9	AV974977
34	64.2	4.0	560	12	BP007496
35	64.2	4.0	585	9	AV965651
36	64.2	4.0	638	13	BW282449
37	64.2	4.0	652	9	AV991679
38	64.2	4.0	656	13	BW276796
39	64.2	4.0	656	13	BW292110
40	64.2	4.0	675	13	BW274166
41	64.2	4.0	675	13	BW302983
42	64.2	4.0	680	13	BW208298
43	64.2	4.0	684	13	BW248944
44	64.2	4.0	685	13	BW251033
45	64.2	4.0	686	13	BW252952

#### ALIGNMENTS

RESULT 1  
AW672518  
LOCUS  
DEFINITION  
LGL\_360\_F03.b1\_A002 Light Grown 1 (LGI) Sorghum bicolor cDNA, mRNA  
588 bp mRNA linear EST 19-JUL-2000  
sequence.  
ACCESSION  
AW672518  
VERSION  
AW672518.1 GI:7536439  
KEYWORDS  
EST.  
SOURCE  
Sorghum bicolor (sorghum)  
ORGANISM  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.  
REFERENCE  
1 (bases 1 to 588)  
AUTHORS  
Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C. and Pratt, L.H.  
TITLE  
An EST database from Sorghum: light-grown seedlings  
JOURNAL  
Unpublished  
COMMENT  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV  
High quality sequence stop: 580  
POLYA-No.

# FEATURES

Location/Qualifiers  
1. .588  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/db\_xref="taxon:4558"  
/clone\_lib="Light Green 1 (LGI)"  
/note="Organ: 10- to 14-day-old light-grown (greenhouse) seedlings; Vector: Lambda zap; Site.1: XhoI; Site.2: EcoRI ; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."  
145 a 151 c 152 g 140 t

BASE COUNT  
ORIGIN

Query Match 8.6%; Score 138.4; DB 9; Length 588;  
Best Local Similarity 59.1%; Pred. No. 3.4e-29;  
Matches 279; Conservative 0; Mismatches 181; Indels 12; Gaps 2;

QY 38 AGTACCCAGAACCCACCGGGTCAAGACTCTTACACCGCAACAATGTCCACATCC 97  
DB 126 ATCTGCCAAGAGAGTACATGACTACAAAGACGGCTACTGCACCAACAATGTCACTATC 185  
QY 98 GTTACAGGAACCCGGGCGAGAGGGCTCTGCGAGACTACCCCGGGGTCAAAATCCCTACT 157  
DB 186 GATACAGAACCC-----TGGTATCTGTGAGAGCAGCCCTGGTGTAGACAGCTACT 236  
QY 158 CTGGATATGTCGACACCTCTCCGAGTCCCATACCTTCTTCTGTGTTCTTCCGAAGCCAGAC 217  
DB 237 CAGGATAGTGGATCTCGCACCAATGTCACGCTCTCTTCTGTGTTCTTCCGAGTCTAGGA 296  
QY 218 ATACCCAGAAATGACCTATCACATCTGTTGATGTGGTGGCCCTGGAGCGATTCTT 277  
DB 297 ACAATCCCGCATCTGATTCCTTCTTCTGTTGATGTGGTGGCCCTGGAGTATTCGC 356  
QY 278 TGATCGGTCTCTTCCGAAGAGTGTGGCCCTTGGCAATGTCATCTTGTGATGACTACA 337  
DB 357 TGATGGGCTCTTGAAGAAACGGTCCATGCATGATTAGTGATAATCTCAGCGCTGAT 416  
QY 338 TCACCCCTACTCTGTGAAGAGGTCTCCAAATTTACTATTCCTGTCCTCCAGCATTTGGAG 397  
DB 417 ACAACCCCTACAGTGAACAATGTCTCCAAACATGCTATATTTCCAGCGGTCTGGCA 476  
QY 398 TCGSCTTTTCATAGTACAGTCTGATGGTCCATTAACCTGTAACCTGGG---TCG 454  
DB 477 CGGATTCAGCTACAGGACAGGCGTCGGCAGTTTGAATCCAGTCACAGGAATTC 536  
QY 455 TCGAATAATTCAGCTTTGACGAGTTCAGGCGCGGTACCCCAACCATTTGATCC 506  
DB 537 GTTACAACTCGAGTGAATACCGGCTGATGGCGGTACCGGCTCTTGGATCC 588

# RESULT 2

CA747625  
LOCUS  
DEFINITION  
GAP13D07 Aspergillus flavus Lambda zap cDNA Library Aspergillus  
flavus cDNA, mRNA sequence.  
CA747625  
VERSION  
KEYWORDS  
SOURCE  
EST.  
GI:27735664

# ORGANISM

Aspergillus flavus  
Aspergillus flavus  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
1 (bases 1 to 721)

# REFERENCE

AUTHORS  
O'Brien, G.R., Fakhoury, A.M. and Payne, G.A.  
TITLE  
Identification of Genes Differentially Expressed During Aflatoxin  
Biosynthesis in Aspergillus flavus and Aspergillus parasiticus  
Unpublished (2003)

# JOURNAL

COMMENT  
Contact: Payne, GA  
Department of Plant Pathology

North Carolina State University  
Box 7616, Gardner Hall, Raleigh, NC 27695, USA  
Tel: (919) 515-6994  
Fax: (919) 515-7716  
Email: Gary\_Payne@NCU.

# FEATURES

Location/Qualifiers  
1. .721  
/organism="Aspergillus flavus"  
/mol\_type="mRNA"  
/strain="3357"  
/db\_xref="taxon:5059"  
/clone\_lib="Aspergillus flavus Lambda zap cDNA Library"  
/note="vector: philesript; Log phase of aflatoxin  
production"  
161 a 159 c 224 g 176 t 1 others

BASE COUNT  
ORIGIN

Query Match 7.2%; Score 116; DB 14; Length 721;  
Best Local Similarity 53.3%; Pred. No. 1.5e-22;  
Matches 245; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

QY 1101 CGGGTCAACATCACTACACCCAGTCCAAATAATGAGCTCTACTACGCTTCCAGCAAC 1160  
DB 4 CGAGGTCCCTGTCACCTTACCGAGCTGCCCGAGTGGTCCACAAACAGTTGACAGAC 63  
QY 1161 AGGGGACTTTGCTGGGCCCACTTCATCGAAGACCTCGAGGAGATCCTTGTCTTCCCGGT 1220  
DB 64 CTAGGACATCACCGCGCGGATTCCTCGACTCGATAGCCACCTCTCGACAGCGGCGT 123  
QY 1221 GCGTGTCTTCTATCTATGCGGAGCGGATTCATCTGCAACTGTTGCGCGTCAGGC 1280  
DB 124 CAGGTACATGATGATGACGCGACCGCTGATGATGATGATGATGATGATGATGAT 183  
QY 1281 CGTTTCCCTGCTCGGAATCTCTCCAAAGCGGCCAGTTCGGAAGCGCAGGATACAGCC 1340  
DB 184 AGCCAGCTCTCCAGTACCTGCTCTCGGCGCGCAGAGTTCGCAACACCGGATATCTCC 243  
QY 1341 CCTGAAAGTCAACGGCGTTCGAGTATGGGGAACCTCGCGAGTATGTTATTTCTTCTTCC 1400  
DB 244 ATTAGTACGCTCCGAGGATCAAGGATGATGACAGCTCACTAGTATATATAGCTTCC 303  
QY 1401 TCGGCTCTATGACGAGCGCATGAATCCCATCTACTACAGCCCATCCCTCCCTGCAATT 1460  
DB 304 GCGGCTTACAGCTGCTCATGAAGTCCCGGCTTATACAGCGGTTGCGGCGTATGAGAT 363  
QY 1461 GTTTAACGGACTATCTTGGTGGATATCGAGGCGCCAGAGAAGATCTGCGCCAG 1520  
DB 364 CTTCATGGTCTAGCTGGATTCGGATATCTTCTAGTGGAGATAGTATACGGGGA 423  
QY 1521 CTACAAGACGAATGGAAGCGCTACAGTACGCATACACAG 1560  
DB 424 GTTTAAGACGAGTGGACCGAGGATCTTGGCATATTAAG 463

# RESULT 3

BM868983  
LOCUS  
DEFINITION  
mgn002xm24f.b Magnaporthe grisea NS Uni-Zap XR Library Magnaporthe  
grisea cDNA clone mgn002xm24 5', mRNA sequence.  
BM868983  
VERSION  
KEYWORDS  
SOURCE  
EST.  
GI:30405680

# ORGANISM

Magnaporthe grisea (anamorph: Pyricularia grisea)  
Magnaporthe grisea  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.  
1 (bases 1 to 498)

# REFERENCE

AUTHORS  
Ebbole, D.J., Yuan, J., Thomas, T.L., Bobrowicz, P., Lu, G., Bhatte, R.  
TITLE  
Expressed sequence tags from the rice blast fungus, Magnaporthe  
grisea  
Unpublished

COMMENT  
On Mar 7, 2002 this sequence version replaced gi:19236665.



Contact: Ebbole DJ  
Department of Plant Pathology & Microbiology  
Texas A&M University  
Peterson Bldg, MS2132, College Station, TX 77843-2132, USA  
Tel: 979 845 4831  
Fax: 979 845 6483  
Email: d-ebbole@tamu.edu

Chromatogram file of this sequence is available, see contact person  
;Best nr hit (April. 22, 2003) gb|EAA29397.1| hypothetical protein  
[Neurospora crassa] 132 2e-30  
PCR Primers  
FORWARD: T3 primer  
BACKWARD: T7 primer  
Plate: mgns002 row: M column: 24  
seq primer: t3.

```

seq primer: 13:
  Location/Qualifiers
    1. .498
      /organism="Magnetaporthe grisea"
      /mol_type="mRNA"
      /strain="Guy11"
      /db_xref="taxon:148305"
      /clone="mgns002xm24"
      /sex="Mat1-2 hermaphrodite"
      /cell_type="mycelium"
      /clone_lib="Magnetaporthe grisea NS Uni-Zap XR Library"
      /note="Vector: pBluescriptSK-; Site_1: EcoRI; Site_2: XhoI;
      /undirectional cloning. EcoRI site has T3 primer and
      predominantly 5' reads. T7 primer on XhoI side of insert.
      Nitrogen starvation library. Cells were inoculated into
      minimal medium and grown for two days with shaking (150
      rpm) at room temperature. Culture was harvested, blended,
      inoculated into minimal medium as above for 24 h. Cells
      were harvested, washed with water and inoculated into
      minimal medium base lacking nitrogen source for 6 h.
      Sequences were processed by one of two methods. Where a
      full-length alignment to the M. grisea genome sequence was
      available, the EST sequence was trimmed according to the
      alignment, otherwise sequence quality was assessed using
      phredPhrap version 99i019 and trimmed according to phd
      files (0.05) and for vector seqs."
    94 a 186 c 143 g 75 t

```

Query Match	5.8%;	Score 92.6;	DB 12;	Length 498;
Best Local Similarity	52.1%;	Pred. No. 8.8e-16;		
Matches 231;	Conservative 0;	Mismatches 209;	Indels 3;	Gaps 1;
QY	1087	GTCTGGAGCGTATCGGCGTCACATCACTACACCCAGTCCATTAATACGCTCTACIAC	1146	
Db	38	GTCCAGCTGGCCCTCGGCGTACCCCTCACTGGACAGCTCTCTCGCGCATGTCGAAG	97	
QY	1147	GCTTTCCAGCAACAGCGCGACTTTGTCTGGCCCAACTTCATCGAAGACCTCGAGGAGATC	1206	
Db	98	GCTTTCGGTGTCATCGGCGACTACAAACGCCCGCGCTGGCTCGAAGCCATGGCAACCTC	157	
QY	1207	CTTGCTCTCCCGTGGGTGTCTTCCTCATCTATTGGCGAGCGGATTCATCTGCAACTGG	1266	
Db	158	CTCGAACAAGGGAAGAAGTGCAGTCTATGTACGGCGACCGGANTTCGCTCGCAACTGG	217	
QY	1267	TTGCGGGTFCAGCGCGTTTCCCTCGTGTGCGAATCTCCCAAGCCGCCAGTTCGGAAGC	1326	
Db	218	ATGGCGGGTGAAGCGCTCTCTTGGCCATTCGGGTGGAGAACCCAGGAAAGTTCGCCGAG	277	
QY	1327	GCAGGGTACACGCCCTTGAAGTCAACAGG---CGTCGAGTATGGGGAACCTCGCGAGTAT	1383	
Db	278	CGGGGTACAGCCCTCTGGGACCACTGCACCTAGACGGCGGGTGTGGCGGCACTAC	337	
QY	1384	GGTAAATTTCTCTCACTCGGCTATGAGGAGCGGCATGAAGTCCCATACTACCACCC	1443	
Db	338	TCCAACCTGACCTTTGCCCGCGTCTTCCAGGCGGACACGCCGACCCCTCGTACCAGCAG	397	
QY	1444	ATCGCCTCCCTGCAATTTGTTAACCGGACATATCTTCGGTTGGATATCCGAGGGGCCAG	1503	

398 CAGACGGCTACCGCATCTTCAACCGGGCCGTCGACAGGACGTGGCGGGCGGCTG 457

1504 AAGAAGATCTGGCCAGCTACAA 1526

458 GTGGACACGGCGAGCAACCCGA 480

RESULT 4  
 Locus CNS0625E 963 bp DNA linear GSS 07-JUL-2001  
 LOCUS T3 end of clone AY0AA015G04 of library AY0AA from strain CBS 6340  
 DEFINITION of Kluyveromyces thermotolerans, genomic survey sequence.

ACCESSION AL421896  
 VERSION AL421896  
 KEYWORDS  
 SOURCE Kluyveromyces thermotolerans  
 ORGANISM Kluyveromyces thermotolerans  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

REFERENCE 1 (bases 1 to 963)  
 AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,  
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,  
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
 Wincker,P. and Weissenbach,J.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)  
 MEDLINE 20584711  
 PUBMED 11152876

REFERENCE 2 (bases 1 to 963)  
 AUTHORS Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P.  
 and Dujon,B.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 10.  
 Kluyveromyces thermotolerans  
 JOURNAL FEBS Lett. 487 (1), 61-65 (2000)  
 MEDLINE 20584720  
 PUBMED 11152885

REFERENCE 3 (bases 1 to 963)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 COMMENT This GSS is part of a random genomic sequencing program of thirteen  
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.

FEATURES  
 source Location/Qualifiers  
 1..963  
 /organism="Kluyveromyces thermotolerans"  
 /mol\_type="genomic DNA"  
 /strain="CBS 6340"  
 /db\_xref="taxon:4916"  
 /clone="AY0AA015G04"  
 /clone\_lib="AY0AA"  
 /note="end : T3"  
 misc\_feature <270..>962  
 /note="similar to Saccharomyces cerevisiae ORF YBR139w [ strong similarity to carboxypeptidase ]"  
 /evidence=not\_experimental

BASE COUNT 242 a 231 c 217 g 272 t 1 others  
 ORIGIN

Query Match 5.4%; Score 86.6; DB 29; Length 963;  
 Best Local similarity 55.5%; Prad. No. 7.7e-14;



Sohngaardsholmsvej 49, 9000 Aalborg, Denmark  
Tel: +45 96358467  
Fax: +45 98141808  
Email: kgw@bio.auc.dk

# FEATURES

source Location/Qualifiers  
1. 696  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/dev\_stage="Plant 3 weeks old, three days post infection"  
/clone\_lib="Infected Arabidopsis Leaf"  
/note="Organ: Leaf; Vector: pBluescript; Mixed cDNA library of Arabidopsis and E. cichoracearum infected leaf from three weeks old Arabidopsis plants. Plants were harvested 3 days after infection and mRNA oligo dt selected."  
BASE COUNT 207 a 129 c 160 g 200 t  
ORIGIN

Query Match 5.2%; Score 83.6; DB 13; Length 696;  
Best Local Similarity 58.5%; Pred. No. 4.8e-13;  
Matches 165; Conservative 0; Mismatches 114; Indels 3; Gaps 1;  
QY 121 GCGCTGCGGAGACTACCCGGGTGTCAAATCCTACTCTGGATATCTCGACACCTCTCC 180  
Db 146 GTATCTGTGAGACGACAGCAGGTGTTAACCAATATCTTGGATATCTTCCGATAGTAT 205  
QY 181 GAGTCCCAACCTCTTCTGGTCTTCGAGAGCAGACATAACCCAGAACTGCACCTATC 240  
Db 206 GATGAGAACATGTGTTCTGGTCTTCGAGGCTAGAAATAATCCAATTAACGGCACCAGTA 265  
QY 241 ACATTGTGTTGAATGGTGGCCCTGGAAGCGATCTTTGATCGTCTCTTCGAGAGTTG 300  
Db 266 GTGGTGTGCTCAATGGGGCCCGCTGCTCAATCAATGATGCTTATTCAGAAAT 325  
QY 301 GGCCTTGGCA--TGTCATTCGACTTTTGATGACTACATCAACCTCACTCGTGGAA 357  
Db 326 GGTCTTGTCAATTTGAAGATGATCCATGAACCATCCCTCAACCCGTACAGCTGGAAT 385  
QY 358 GAGGTCTCAATTTACTATTCCTGCCAGCCATGGGAGTC 399  
Db 386 GAATTGGCCAAACGCTACATTTGACCAAGAGCTTCAAAATC 427

RESULT 7  
BI200653  
LOCUS  
DEFINITION  
BI200653 485 bp mRNA linear EST 10-JUL-2001  
ola06fs.r1 Fusarium sporotrichioides Tri 10 overexpressed cDNA library Fusarium sporotrichioides cDNA clone ola06fs 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Fusarium sporotrichioides  
Fusarium sporotrichioides  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Mitosporic Hypocreales; Fusarium. 1 (bases 1 to 485)  
Ren,Q., Tag,A., Peplow,A., Lai,H., Kupfer,C., Peterson,A., Beremand ,M. and Roe,B.  
Analysis of a Fusarium sporotrichioides EST database  
Unpublished  
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu  
Department of Chemistry and Biochemistry  
Advanced Center for Genome Technology, University of Oklahoma  
620 Parrington Oval, Norman, OK 73019, USA  
Tel: 405 325 4912  
Fax: 405 325 7762  
Email: broe@ou.edu  
Contact Dr. Marian Beremand regarding clone availability Included is the best homolog from a blastx search of Genbank nr 04-09-01 466 2e-46 gi|461830|sp|P34946| CARBOXYPEPTIDASE Slpir||S38953

carboxypept  
Seq primer: T3  
High quality sequence stop: 467.  
Location/Qualifiers  
1. 485  
/organism="Fusarium sporotrichioides"  
/mol\_type="mRNA"  
/strain="Tri 10"  
/db\_xref="taxon:5514"  
/clone="ola06fs"  
/clone\_lib="Fusarium sporotrichioides Tri 10 overexpressed cDNA library"  
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript ; 3' end of cDNA cloned into XhoI site of pBluescript"  
BASE COUNT 122 a 128 c 120 g 115 t  
ORIGIN

Query Match 5.0%; Score 80.6; DB 12; Length 485;  
Best Local Similarity 57.5%; Pred. No. 2.9e-12;  
Matches 165; Conservative 0; Mismatches 119; Indels 3; Gaps 1;  
QY 121 GCGCTGCGGAGACTACCCGGGTGTCAAATCCTACTCTGGATATCTCGACACCTCTCC 180  
Db 177 GGAATATGCGAGACTACGCCAGGTGTCGGCCAGCACTCTGTTTCTCTGTCGGGAC 236  
QY 181 GAGTCCCAACCTCTTCTGGTCTTCGAGAGCAGACATAACCCAGAACTGCACCTATC 240  
Db 237 AACATGAACATGTTCTGGTCTTCGAGGCTAGAAATAATCCAATTAACGGCACCAGTA 296  
QY 241 ACATTGTGTTGAATGGTGGCCCTGGAAGCGATCTTTGATCGTCTCTTCGAGAGTTG 300  
Db 297 GCCTTGGCTCAATGGGGCCCGCTGCTCAATCAATGATGCTTATTCAGAAAT 356  
QY 301 GGCCTTGGCAATCAATTCG--ACATTGATGACTACATCAACCTCACTCGTGGAA 357  
Db 357 GGCCTTGGCAATCAATTCG--ACATTGATGACTACATCAACCTCACTCGTGGAA 416  
QY 358 GAGGTCTCAATTTACTATTCCTGCCAGCCATGGGAGTCGCTT 404  
Db 417 ACATTGGCCAAACGCTACATTTGATGATGATGATGATGATGATGATGATGATGATGAT 463

RESULT 8  
BI187544  
LOCUS  
DEFINITION  
BI187544 445 bp mRNA linear EST 10-JUL-2001  
a4e08fs.r1 Fusarium sporotrichioides Tri 10 overexpressed cDNA library Fusarium sporotrichioides cDNA clone a4e08fs 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Fusarium sporotrichioides  
Fusarium sporotrichioides  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Mitosporic Hypocreales; Fusarium. 1 (bases 1 to 445)  
Ren,Q., Tag,A., Peplow,A., Lai,H., Kupfer,C., Peterson,A., Beremand ,M. and Roe,B.  
Analysis of a Fusarium sporotrichioides EST database  
Unpublished  
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu  
Department of Chemistry and Biochemistry  
Advanced Center for Genome Technology, University of Oklahoma  
620 Parrington Oval, Norman, OK 73019, USA  
Tel: 405 325 4912  
Fax: 405 325 7762  
Email: broe@ou.edu  
Contact Dr. Marian Beremand regarding clone availability Included is the best homolog from a blastx search of Genbank nr 04-09-01 386 5e-37 gi|461830|sp|P34946| CARBOXYPEPTIDASE Slpir||S38953  
carboxypept  
Seq primer: T3

[illegible]

exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

# FEATURES

Location/Qualifiers  
source

1..1054  
/organism="Saccharomyces kluyveri"  
/mol\_type="genomic DNA"  
/strain="CBS 3082"  
/db\_xref="taxon:4934"  
/clone\_lib="AU0A015E08"  
/clone\_lib="AU0A0A"  
/note="end : 17"

## misc\_feature

<148..>1052  
/note="similar to Saccharomyces cerevisiae ORF YMR297w [pCl : carboxypeptidase Y, serine-type protease]  
2 putative frameshift(s)  
/evidence-not\_experimental"

BASE COUNT 272 a 219 c 229 g 329 t 5 others

## ORIGIN

Query Match 4.8%; Score 77; DB 29; Length 1054;  
Best Local Similarity 53.1%; Pred. No. 5.4e-11;  
Matches 164; Conservative 0; Mismatches 145; Indels 0; Gaps 0;  
QY 187 CATACCTTCTTGTGTTCTCGAAGCCAGACATAACCCAGAACTCAGCTATCACATIG 246  
DB 286 CATTTCTTCTACTGGTTTITGAAGTAGGAATGACCCAAAATACACCTAICGTTCIT 345  
QY 247 TGGTTCAATGGGCGCTCGAAGCGATTCTTTGATCGGTCTCTTCGAAGAGTTGGCCCT 306  
DB 346 TGGTTGAAGCGAGGTCAGGTTGCTCCTCTTTGACTGGTTTATTTTCGAATTTGGTTCT 405  
QY 307 TGCCATGTCAATTCGACTTTTGTGACTACTACATCAACCTTCACCTGTGGAGCGAGGCTCC 366  
DB 406 TCTTCAATTTGGCCAGATTGTAACCAATTTACACCCCTTACCTTGGATTCCCAATGCC 455  
QY 367 AATTACTATTCCTGTCACGCAATGGGAGTGGGCTTTTCATATAGTATAGCGTTGAT 426  
DB 466 TCGGTCAICTTTTGGACCAAGCAGTTAAGTCGGTTATTCCTACCTTCCTCTGAAGGT 525  
QY 427 GGTCCATTAAACCTTACTACTGGGTCTCGAATAATTCAGCTTTCAGAGTTCAGGC 486  
DB 526 GTTTCAGACACCGTTGCTCGCGCAAGGATGATATGCTTCTTCGAGCTGTTTTCAGA 585  
QY 487 CGGTACCCA 495  
DB 586 CAGTCCCA 594

## RESULT 11

BQ110016  
LOCUS BQ110016 729 bp mRNA linear EST 01-APR-2003  
DEFINITION VD0102006 VD01 Verticillium dahliae cDNA, mRNA sequence.  
ACCESSION BQ110016  
VERSION BQ110016.1 GI:29426386  
KEYWORDS EST.

## SOURCE

Verticillium dahliae  
Verticillium dahliae  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreales; Hypocreales; mitosporic Hypocreales;  
1 (bases 1 to 729)

## REFERENCE

Neumann,M.J. and Dobinson,K.F.  
Sequence tag analysis of gene expression during pathogenic growth and microscletoria development in the vascular wilt pathogen Verticillium dahliae  
Fungal Genet. Biol. 38 (1), 54-62 (2003)  
22442000

# PUBMED

12553936

Contact: Dobinson KF  
Agriculture & Agri-Food Canada  
1391 Sandford St., London, Canada, N5V 4T3  
Email: dobinsonk@agr.ca  
Plate: 2 row: 67 column: 6.

## FEATURES

source

1..729

/organism="Verticillium dahliae"  
/mol\_type="mRNA"  
/strain="Dvd-T5"  
/db\_xref="taxon:27337"  
/lab\_host="E. coli"  
/clone\_lib="VD01"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Developing microscletoria (DMS) cDNA library; constructed from axenic cultures grown on cellulose membranes overlaid onto a basal medium agar. Cells were harvested at 4 days post-inoculation. Single-pass sequencing was done using the T3 promoter primer: 5' ATATACCCCTACATAAGGA 3'."

BASE COUNT 152 a 243 c 188 g 144 t 2 others

## ORIGIN

Query Match 4.6%; Score 74; DB 13; Length 729;  
Best Local Similarity 58.7%; Pred. No. 3.3e-10;  
Matches 128; Conservative 0; Mismatches 90; Indels 0; Gaps 0;  
QY 193 TTCTTCGTCTTCGAGCCAGACATACCCAGAACTCAGCTATCACATTTGSGTTG 252  
DB 171 TTCTACTGTTCTTCGAGTCCGCAATGACCCCAAGAACGCCGCTCTCTGSGTCTC 230  
QY 253 AATGGTGGCCCGAGAGCGATTCTTTGATCGGTCTCTTCGAAGAGTTGGCCCTTGCCAT 312  
DB 231 AACGGTGGCCGCTGTTGCTCTTCCTCAGCGGCTCTTCATGAGTTCGCGCCGCGCAGC 290  
QY 313 GTCAATTCGACTTTTGTGACTACTACATCAACCTTCACCTGTGGAGAGGTTCCCAATTA 372  
DB 291 ATCAACAAGAGCTCGAGATGTTCACACAGAGTGGTCTTGGACACACACGCTTCGGTC 350  
QY 373 CTATTCCTGTCGCCAGCATTTGGAGTCGGCTTTTCATA 410  
DB 351 ATCTTCCTTGACCAAGCTGTGAACTGGTCTACTCTGTA 388

## RESULT 12

CNS0620K/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED



```

CD054523
VERSION CD054523.1 GI:30595544
KEYWORDS EST,
SOURCE Hordeum vulgare
ORGANISM Hordeum vulgare

REFERENCE
1 (bases 1 to 682)
Zierold, G. and Schwoizer, P.
Barley ESTs from pathogen-attacked leaf epidermis
Unpublished
Contact: Patrick Schweizer
Transcriptome Analysis, Cytoogenetics Department
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, D-06466 Gatersleben, Germany
Tel: 0049 (0)39482-5660
Fax: 0049 (0)39482-5595
Email: schweiz@ipk-gatersleben.de
Insert Length: 682 Std Error: 0.00
Plate: 1 row; B column: 22
Seq primer: M13rev.
Location/Qualifiers
1..682
/organism="Hordeum vulgare"
/mol_type="mRNA"
/cultivar="Ingriid BC mlo-5"
/db_xref="GABI:703742"
/db_xref="taxon:4513"
/clone="H001B22"
/tissue_type="leaf epidermis, 6 h and 24 h post
inoculation with Blumeria graminis"
/dev_stage="7 d after germination"
/lab_host="XL10-Gold"
/clone_lib="HO"
/notes="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
cDNA); Site_2: XhoI (3'-end of cDNA); Approximately 5 % of
the clones correspond to cDNA from the fungi B. graminis
hordel and tritici, respectively. Due to a cloning
artefact caused by the kit, in most cases the EcoRI site
is NOT present, as well as the EcoRI adapter used for
cloning. To excise the insert, restriction sites upstream
EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
due to the cloning system used Blue/white selection for
recombinants is not 100% reliable. Average insert size is
1.2 kb."
BASE COUNT 191 a 173 c 152 g 166 t
ORIGIN
Query Match 4.5%; Score 72.2; DB 14; Length 682;
Best Local Similarity 57.2%; Pred. No. 1.1e-09;
Matches 131; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 187 CATACCTTCTCTGGTCTTCGAGCCGACATACCCAGAACTGCACCTATCACATTG 246
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 CATCTTTTACTGGTTTTTTGAGTCCGAAACGATCCGAGAGACCCAGTCATCTG 497

QY 247 TGGTTGATGTGGCCCTGGAGGCAATCTTTGATCGGTCTCTTCGAGAGATTGGCCCT 306
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 498 TGGCTCAACGCGGCCCTGGAGTGTTCGTCATTAACGATTAATCTTTGAGCTCGGCCG 557

QY 307 TGGCATGTCAATTCGACTTTTGATGACTTACATCAACCTCACTCGTGAACGAGGTGCC 366
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 GCCTCAATCGATAAAACTTGAAGTGTGTCATACCTTATCTTGGAACTCGAACGCA 617

QY 367 AATTACATTCCTGTCGCCGCAATGGAGTGGCTTTTCATAPAGT 415
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 618 TCAGTTATTTTCCCTGACCAACCTGTTAAIGTCGATACGTCATACGCG 666

RESULT 15
CD456151/c
LOCUS CD456151
DEFINITION Fg03_10g07.R
Fg03_AaFC_ECORC_Fusarium_graminearum_mycelium_trichothecene_product
ion Gibberella zeae cDNA clone Fg03_10g07, mRNA sequence.
ACCESSION CD456151
VERSION CD456151.1 GI:31370891
KEYWORDS EST,
SOURCE Gibberella zeae
ORGANISM Gibberella zeae

REFERENCE
1 (bases 1 to 907)
Harris, L.J., Allard, S., Rochleau, H., Chapados, J., Couroux, P.,
Hattori, J., Masotti, M., Ouellet, T., Robert, L.S., Singh, J.A., Sprott
, D. and Finker, N.A.
A cDNA library prepared from Fusarium graminearum grown in liquid
culture under trichothecene-producing conditions
Unpublished
Contact: Harris, Linda J.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
CANADA
Tel: (613) 759-1314
Fax: (613) 759-6566
Email: harris.lj@em.agr.ca.
Location/Qualifiers
1..907
/organism="Gibberella zeae"
/mol_type="mRNA"
/strain="DAOM 180378"
/db_xref="taxon:5518"
/clone="Fg03_10g07"
/tissue_type="Mycelium"
/dev_stage="Asexual"
/lab_host="E. coli SURE2"
/clone_lib="Fg03_AaFC_ECORC_Fusarium_graminearum_mycelium_
trichothecene_production"
/notes="Vector: pGEM-T easy; Site_1: EcoRI; Fusarium
graminearum grown in liquid culture to optimize
trichothecene production; final deoxynivalenol
concentration in culture determined by CD-ELISA to be 165
micrograms/ml. cDNA was prepared using a modified SSH
protocol to promote normalization whereby cDNA was not
digested prior to hybridization and no subtraction was
attempted."
BASE COUNT 243 a 192 c 225 g 242 t 5 others
ORIGIN
Query Match 4.4%; Score 70.2; DB 14; Length 907;
Best Local Similarity 45.9%; Pred. No. 4.9e-09;
Matches 379; Conservative 2; Mismatches 410; Indels 34; Gaps 3;

QY 670 TTTACGACGAGATGAGAGATTGCCAACGGTAGTGTATTAATGGTTCAGCTTAATTTC 729
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 906 TTTGAACGACGAGATGCTGCTATKATGCTGTACCATCAAGGCGGAGAGATCAACSTC 847

QY 730 AACTCTCTGGGATTTATTACGGCATATCGACGAGGATCCAGGCCCTTACTACCTT 789
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 846 GTAGCTCTGGCATCAACAATGGATGATCTCTGCCAACACGATACAGATTACATC 787

QY 790 GAATTCGCTGCAACAATACCTACGCTATCAAGGCTGTCAACGAGACCGCTTACAACTAC 849
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 796 GAATACGCGCAATACACCTACAAAAGYGTGATCACTCACTCACTACAGCAATAC 727

QY 850 ATGAGATTGGCCAAACAAATGCGAGATTTGCCAGATTTGATTCACCTTCCAAAGAC 909
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 726 CTAACACCTACATCAAAAG-----TGTTGGCTGTTTCGCS 689

QY 910 ACAAAACCCAGGATTAAGCTAGCTACGCCCTCTCGCGGAGCCACCAACATGTGAGG 969
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 688 AAGTGCCTCCAGGCTCACAGGCAACGACGCTGTTGTGGCAATGCCGAGATGTCGTGAGI 629

QY 970 GACAAATGTTAGGGGCCCAIACCTACGCTTGTCTGCTGCTGTTGTTGTTATATCGGCAT 1029

```

Db	628	CGCGTATTGAACGGCCCTTGGAGATACAGCAGACTTTGACGTTTATGATATTCGGGC	569
Qy	1030	CCATATGATGACCCGACTCGCCAGCTTATTACACAAATTTCTGGCAAGGACTCTGTC	1089
Db	568	CCGAGTAAGCACCCTTTTCCTCATCGACTTACTCGACTTAUCTTCAGTCAATCAAGTGC	509
Qy	1090	ATGGAGCGCTATCGGCGTCAACATCACTACACCCAGTCCAAATATGACGTCFACCTACGCT	1149
Db	508	ATGAAAGCTATTGGCGCGAGTCAAGATACGGCGAATGTCCGAGGCGGTTATGACAAG	449
Qy	1150	TTCACGAAACAGCGACTTTGTCTGGCCAACTTCATCGAAGACCTCGAGGAGATCCCT	1209
Db	448	TTTATCAACAGTGGCGATCGGGAACGATCATTTCTGTCACATGTTCCAAAG--TCATT	392
Qy	1210	GCTCTCCCGTGGGTGTCCTCATCTATGGCGACGCGATTACATCTGCAACTGGTTC	1269
Db	391	GACTCCAAGATCAAGTATGATCTGGCTGGTGGATGCCGACTGGATCTGCAACTGGAT-	333
Qy	1270	GGCGGTTCAGGCGTTCCTTCGCTGGCGAATCTCCAGCGCCGCAAGTTCGGAACGCA	1329
Db	332	-----GGGTAAATACCGCGCTCTCAACTATTGCTCCCGACGTCTGCTCTCAGCT	281
Qy	1330	GGGTACAGCGCCCTGAAATCAACGCGTCGAGTATGGGAACTCCGAGTAIGTAAAT	1389
Db	280	CCTCTTCAATCCCTTACTGTTGATGGAACHTAAGTACGGAGATTTAAGACATCTGGAAAT	221
Qy	1390	TTCTCCTTCACTCGCCTCTATGAGGCGAGGCCATGAAGTCCCATACTACCGACCAATCGGC	1449
Db	220	TTGAGCTGGCTGGAGTTATGGTGTGTGTCATGAGGTCCCSCTTATCAGCCTCAGGCT	161
Qy	1450	TCCCTGGAATGTTTAACGGGACTATCTCGGTTGGGATATCGCA	1494
Db	160	GCATTGGCTGCTTTGTGCGCCACATTTGCCAAGAGCCTATTCA	116

Search completed: September 16, 2003, 15:55:38  
Job time : 4190 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 21:25:11 ; Search time 74 seconds  
(without alignments)  
1151.840 Million cell updates/sec

Title: US-09-712-338-2\_COPY\_19\_555

Perfect score: 2887

Sequence: 1 LPGSTPASVGRRLPKNPITG.....HTQSSVPLTATSMSSVGMGA 537

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

1:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2883	99.9	554	19	AAW56099
2	2338	81.0	551	23	ABR38864
3	980.5	34.0	623	23	ABR38865
4	941	32.6	554	23	ABR38859
5	916.5	31.7	526	23	ABR38817
6	636	22.0	508	23	ABR381049
7	630	21.8	536	23	ABR38819
8	630	21.8	556	17	AAR96737
9	620	21.5	557	17	AAR96738
10	554	19	AAW56099		A. oryzae ATCC2038
11	551	23	ABR38864		A. niger serine ca
12	623	23	ABR38865		A. niger serine ca
13	554	23	ABR38859		A. niger serine ca
14	526	23	ABR38817		A. niger serine ca
15	508	23	ABR381049		S. cerevisiae BAX-
16	536	23	ABR38819		A. niger carboxype
17	556	17	AAR96737		A. niger Bo-1 carb
18	557	17	AAR96738		A. niger SFAG 2 ca

10	568.5	19.7	491	15	AAW48059	Sequence of protea
11	541.5	18.8	550	23	ABG93281	C. albicans BAX-as
12	414	14.3	481	23	ABR38845	A. niger carboxype
13	383	13.3	482	21	AAG30065	Arabidopsis thalia
14	383	13.3	502	21	AAG30064	Arabidopsis thalia
15	365	12.6	455	23	ABR38843	A. niger carboxype
16	361.5	12.5	479	21	AAG23905	Arabidopsis thalia
17	361.5	12.5	486	21	AAG23904	Arabidopsis thalia
18	361	12.5	476	19	AAW72966	Human serine carbo
19	361	12.5	476	22	AAW72966	Human serine carbo
20	359	12.4	476	20	AAW13372	Human membrae or
21	359	12.4	476	21	AAW13372	Human membrae or
22	359	12.4	476	22	AAW29228	Human TANGO 176.
23	359	12.4	476	22	AAW29228	Human PRO polypt
24	359	12.4	476	22	AAW29228	Human hydrophobic
25	359	12.4	476	22	AAW29228	Human PRO223 prote
26	359	12.4	476	24	ABU69650	Novel human secret
27	359	12.4	476	24	ABU69650	Human secreted/tra
28	359	12.4	476	24	ABU71473	Human PRO223 prote
29	359	12.4	476	24	ABU71473	Human PRO polypt
30	359	12.4	476	24	ABU65773	Human secreted/tra
31	359	12.4	476	24	ABU65773	Human secreted/tra
32	359	12.4	476	24	ABU67610	Novel human secret
33	359	12.4	476	24	ABU67610	Human secreted pro
34	359	12.4	476	24	ABU64527	Human secreted/tra
35	359	12.4	476	24	ABU64527	Human secreted/tra
36	359	12.4	476	24	ABU58604	Human PRO polypt
37	359	12.4	476	24	ABU58604	Human secreted/tra
38	359	12.4	476	24	ABU57135	Human PRO polypt
39	359	12.4	476	24	ABU54375	Human secreted/tra
40	359	12.4	476	24	ABU10714	Human secreted/tra
41	357.5	12.4	477	22	AAW25810	Human protein sequ
42	357.5	12.4	479	21	AAG47178	Arabidopsis thalia
43	357	12.4	486	21	AAG47177	Arabidopsis thalia
44	357	12.4	476	20	AAW28570	Secreted peptide C
45	357	12.4	476	22	AAU39043	Human secreted pro
						Human polypeptide

## ALIGNMENTS

## RESULT 1

AAW56099  
ID AAW56099 standard; Protein: 554 AA.

XX AAW56099;

AC AAW56099;

DT 27-AUG-1998 (first entry)

XX A. oryzae ATCC20386 carboxypeptidase I protein.

XX Carboxypeptidase I; flavour improving agent; hydrolysate; proteinaceous; food industry.

XX Aspergillus oryzae.

XX Key Location/Qualifiers

FT Peptide 1..18

FT Protein /label= signal

FT Protein /label= carboxypeptidase I

XX WO9814599-A1.

XX 09-APR-1998.

XX 03-OCT-1997; 97WO-US17977.

XX 27-NOV-1996; 96US-0757534.

XX 04-OCT-1996; 96US-0726880.

XX (NOVO) NOVO NORDISK BIOTECH INC.

XX (NOVO) NOVO-NORDISK AS.

APP. 102b

102e

XX Berka R, Blinkovsky A, Brown K, Danbmann C, Golightly E;  
PI Klotz A, Mathisen TE, Rey M;  
XX WPI; 1998-240098/21.  
DR N-PSDB; AAV28620.  
XX Carboxypeptidase from *Aspergillus oryzae* - hydrolyses proteinaceous  
PT substrates, useful for improving flavour of foods  
XX Claim 1; Fig 3; 82pp; English.  
XX This sequence represents carboxypeptidase I from *Aspergillus oryzae*.  
CC This polypeptide has an optimal activity in the range of pH 3.0-7.5 at  
CC 25 deg. C, optimal activity in the range of 55-60 deg. C at pH 4, and a  
CC residual activity of at least 65.5% after 30 minutes at pH 4.0 and  
CC 60 deg. C. It also has the capacity to hydrolyse X from N-Cbz-Ala-X where  
CC N-Cbz is N-carboxybenzoyl and X is any amino acid. The carboxypeptidases  
CC can be used for obtaining hydrolysates (which can be enriched in free  
CC glutamic acid or peptide bound glutamic acid residues) from proteinaceous  
CC substrates. The carboxypeptidases can be used in flavour-improving  
CC compositions in the food industry. The products can also be used for the  
CC production of polypeptides free of carboxypeptidase activity.  
XX Sequence 554 AA;  
XX  
XX Query Match 99.9%; Score 2883; DB 19; Length 554;  
XX Best Local Similarity 100.0%; Pred. No. 1.9e-244;  
XX Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LPGSTASVGRRLPKNPTGVKTLTANNVTIRYKEPGAEGVCTTEPGVKSYSYVDTPSP 60  
Db 19 LPGSTASVGRRLPKNPTGVKTLTANNVTIRYKEPGAEGVCTTEPGVKSYSYVDTPSP 78  
QY 61 ESHTFFFFFAHNPETAPITLWNGPGSDSLGLFEELGPGCHVNSTFDYINPHSWNE 120  
Db 79 ESHTFFFFFAHNPETAPITLWNGPGSDSLGLFEELGPGCHVNSTFDYINPHSWNE 138  
QY 121 VSNLLFLSQPLGVGSYSDVVGSIINPTGVVNSSFAGVQGYPTDAILDTNLAEE 180  
Db 139 VSNLLFLSQPLGVGSYSDVVGSIINPTGVVNSSFAGVQGYPTDAILDTNLAEE 198  
QY 181 AAEWILQGLSLPSLRVSKDFSLWTSYGGHYGPAFNFHFEQNERIANGSVANGVQ 240  
Db 199 AAEWILQGLSLPSLRVSKDFSLWTSYGGHYGPAFNFHFEQNERIANGSVANGVQ 258  
QY 241 LNFNSLGIINGIIDEAIOAPYYPFAVNNIYGIKAVNETVYNNKMFANQMPNGCQDLIST 300  
Db 259 LNFNSLGIINGIIDEAIOAPYYPFAVNNIYGIKAVNETVYNNKMFANQMPNGCQDLIST 318  
QY 301 CKQTNRTALADYALCAEATNMCNVEGPPYAFAGRGYDIDRHPYDDPTPPSYNKKFLAK 360  
Db 319 CKQTNRTALADYALCAEATNMCNVEGPPYAFAGRGYDIDRHPYDDPTPPSYNKKFLAK 378  
QY 361 DSVMDAIGVNIYQSNNDVYAFQQTGDFWPNFIEDLEILALPVKRVSLIYGDAVYIC 420  
Db 379 DSVMDAIGVNIYQSNNDVYAFQQTGDFWPNFIEDLEILALPVKRVSLIYGDAVYIC 438  
QY 421 NWFGGQAVSLAANTYQAAQFRSAGYTLKVNVEYGETREYGNFSFIRVYAGHEVPYQ 480  
Db 439 NWFGGQAVSLAANTYQAAQFRSAGYTLKVNVEYGETREYGNFSFIRVYAGHEVPYQ 498  
QY 481 PIASLQFNRTIFGWDIAEGOKKIWPYSKTKTGATATHTQSSVPLPTATSMSSVGM 536  
Db 499 PIASLQFNRTIFGWDIAEGOKKIWPYSKTKTGATATHTQSSVPLPTATSMSSVGM 554  
XX  
XX RESULT 2  
XX ABR38864  
XX ABR38864 standard; Protein; 551 AA.  
XX  
XX ABR38864;  
XX

DT 24-APR-2003 (first entry)  
XX  
XX DE  
XX XX  
XX A. niger serine carboxypeptidase polypeptide #3.  
KW Protease; fungal infection; aspergillosis; food; tanning; detergent;  
KW protein solubility; viscosity; taste; texture; nutritional value;  
KW EC3.4.16.6.  
XX  
XX *Aspergillus niger*.  
XX OS  
XX WO200268623-A2.  
XX  
XX 06-SEP-2002.  
XX  
XX 22-FEB-2002; 2002WO-EP01984.  
XX  
XX 23-FEB-2001; 2001EP-0200657.  
XX 23-FEB-2001; 2001EP-0200658.  
XX 23-FEB-2001; 2001EP-0200660.  
XX 26-FEB-2001; 2001EP-0200706.  
XX 26-FEB-2001; 2001EP-0200707.  
XX 26-FEB-2001; 2001EP-0200708.  
XX 26-FEB-2001; 2001EP-0200719.  
XX 28-MAR-2001; 2001EP-0000075.  
XX 28-MAR-2001; 2001EP-0000078.  
XX 28-MAR-2001; 2001EP-0000080.  
XX 28-MAR-2001; 2001EP-0000087.  
XX 28-MAR-2001; 2001EP-0000088.  
XX 21-MAY-2001; 2001EP-0000156.  
XX 21-MAY-2001; 2001EP-0000159.  
XX 21-MAY-2001; 2001EP-0000160.  
XX 21-MAY-2001; 2001EP-0000162.  
XX 21-MAY-2001; 2001EP-0000165.  
XX 21-MAY-2001; 2001EP-0000166.  
XX 21-MAY-2001; 2001EP-0000168.  
XX 21-JUN-2001; 2001EP-0000240.  
XX 21-JUN-2001; 2001EP-0000242.  
XX 21-JUN-2001; 2001EP-0000244.  
XX 21-JUN-2001; 2001EP-0000246.  
XX 12-JUL-2001; 2001EP-0000280.  
XX 12-JUL-2001; 2001EP-0000285.  
XX 30-JUL-2001; 2001EP-0000323.  
XX 30-JUL-2001; 2001EP-0000327.  
XX 02-AUG-2001; 2001EP-0000341.  
XX 02-AUG-2001; 2001EP-0000342.  
XX 02-AUG-2001; 2001EP-0000343.  
XX 02-AUG-2001; 2001EP-0000344.  
XX 09-AUG-2001; 2001EP-0000357.  
XX 16-AUG-2001; 2001EP-0000374.  
XX 16-AUG-2001; 2001EP-0000377.  
XX 20-SEP-2001; 2001EP-0000478.  
XX 20-SEP-2001; 2001EP-0000483.  
XX 22-OCT-2001; 2001EP-0000552.  
XX 22-OCT-2001; 2001EP-0000553.  
XX 22-OCT-2001; 2001EP-0000554.  
XX 22-OCT-2001; 2001EP-0000556.  
XX 22-OCT-2001; 2001EP-0000557.  
XX 22-OCT-2001; 2001EP-0000558.  
XX 15-NOV-2001; 2001EP-00004464.  
XX 21-DEC-2001; 2001EP-0005117.  
XX (STAM ) DSM NV.  
XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;  
PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;  
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;  
PI Stiebler J, Albarg R;  
XX WPI; 2002-723203/78.  
XX P-PSDB; ABZ78231, ABZ78288.  
XX  
XX Novel isolated protease polypeptide useful in laboratory, clinical,  
XX pharmaceutical, chemical, diagnostic, personal care and industrial  
XX



polypeptide or polynucleotide of the invention is useful for diagnosing a fungal infection such as aspergillosis, or as a query sequence to perform a search against public databases. A polypeptide of the invention is useful in a selected number of industrial or pharmaceutical processes, in laboratory or clinical processes, in food industry (baking, brewing, cheese manufacturing, meat tenderising), in tanning industry and in the manufacture of biological detergents. A polypeptide may also be useful for improving protein solubility, extraction yields, viscosity or taste, texture, nutritional value, minimising of antigenicity or anti-nutritional factors, colour or functionality as well as processing aspects like filterability of the proteinaceous raw material. The sequences shown in ABR3813-ABR3869 represent the A. niger proteases of the invention.

XX SQ Sequence 623 AA;

Query Match 34.0%; Score 980.5; DB 23; Length 623;  
Best Local Similarity 39.8%; Pred. No. 4e-77;  
Matches 216; Conservative 91; Mismatches 178; Indels 61; Gaps 19;

QY 26 TANNVTIRYKE--PGAE-----GVCETTPGVKSYGVDTSPES----- 62  
DB 32 TPDLLVISEIFPGARISYKQPLGICTTPTSPSYSGYIHLPPHILTNLSIRGISISQ 91  
QY 63 ---HTFFWFPEAR--HNPETAPITELWNGSGSDSLIGLFEELGPCHVNSTFDD--YINPH 116  
DB 92 YPINTFFWPPSRHHNNHNDTSPLTIWNGSGPGSSMIGLFGQENGPCVTNEDSKSTAYNPW 151  
QY 117 SNNEVSNLLFSLPGLGVGSYSTVGSINPVTGVVENSFAGVQRYPIID--ALIDIT 174  
DB 152 SNWEYDMYLIIEQVQTGFSGYDLVRNGTLD-----LNETH--LVGTLFSDQVHGVRNGT 203  
QY 175 TNLAAEAAWELLQGLFSG-----LPSLDSRVQSKD--FSLWTSYGGHGYPAFFNHFYBONE 229  
DB 204 VN--GGRALLVALQVWIGFSEYSSVDNGCGGDDRVSIWTSYGGYGRAYTALFQEMNE 262  
QY 230 RIANGSVN--GVQLNFSNLGILGIDETAIQAPYYPEFAVNNTYGIKAVNETVYNYMKFAN 288  
DB 263 RIESGEVSTGKTHLDTLGIANGCVLLVQVPSFEQAYNNTYVIGIEGINTLYDRAMDSW 322  
QY 289 QMENGCDLSTCKQINRTALAD--YALCAEATNMCRDNVGEPYAFAGRGVYDIRHPYD 347  
DB 323 SKPGGGRDMTIECKRAGE--LGDLPLICEASDYCSREIKSLYNTISGRGYDYDIAHTPD 380  
QY 348 PTPPSYNKLAKSDVDAINVTQSNNDVYAFQQTGDFWVP-----NFIEDLEEI 402  
DB 381 AALVPYFVGFNLRPWQKALGVVNYTMSSEAVGNSFASTGDY--PRNDPRGMIGDIGYL 438  
QY 403 LALPVRVSLIYGADYICNWFQCGQAVSLAANYSCAAQFRSAGYTPPLKVNQVEY--GETRE 450  
DB 439 LDSGVKVMYIGDRDYACPRWGDEYSLIWEYDAEKFRAGYAEVQTKS--SYVGGILVRQ 457  
QY 461 YGNFSTRVYAGHEVPTYOPIASLOLFNRTIFGWDIAEGQKKIW--PSYKTINGTATAT 518  
DB 498 YGNFSTRVYQAGHEVFFYQEPETAYEIFNRAQFNWDIAIGGISLEQNQSYGEGSPSTWH 557  
QY 519 TQSSVP 524  
DB 558 IKNEVP 563

RESULT 4

ABR38859  
ID ABR38859 standard; Protein: 554 AA.

XX AC ABR38859;

XX DT 24-APR-2003 (first entry)

XX A. niger serine carboxypeptidase polypeptide #2.

XX Protease; fungal infection; aspergillosis; food; tanning; detergent;  
KW protein solubility; viscosity; taste; texture; nutritional value;

KW EC3.4.16.6.  
XX Aspergillus niger.  
XX WO200268623-A2.  
PD 06-SEP-2002.  
XX 22-FEB-2002; 2002WO-EP01984.  
PF 23-FEB-2001; 2001EP-0200657.  
XX 23-FEB-2001; 2001EP-0200658.  
PR 23-FEB-2001; 2001EP-0200660.  
PR 26-FEB-2001; 2001EP-0200706.  
PR 26-FEB-2001; 2001EP-0200707.  
PR 26-FEB-2001; 2001EP-0200708.  
PR 26-FEB-2001; 2001EP-0200719.  
PR 28-MAR-2001; 2001EP-0000075.  
PR 28-MAR-2001; 2001EP-0000078.  
PR 28-MAR-2001; 2001EP-0000080.  
PR 28-MAR-2001; 2001EP-0000087.  
PR 28-MAR-2001; 2001EP-0000088.  
PR 21-MAY-2001; 2001EP-0000156.  
PR 21-MAY-2001; 2001EP-0000159.  
PR 21-MAY-2001; 2001EP-0000160.  
PR 21-MAY-2001; 2001EP-0000162.  
PR 21-MAY-2001; 2001EP-0000165.  
PR 21-MAY-2001; 2001EP-0000166.  
PR 21-MAY-2001; 2001EP-0000168.  
PR 21-JUN-2001; 2001EP-0000240.  
PR 21-JUN-2001; 2001EP-0000242.  
PR 21-JUN-2001; 2001EP-0000244.  
PR 12-JUL-2001; 2001EP-0000246.  
PR 12-JUL-2001; 2001EP-0000280.  
PR 12-JUL-2001; 2001EP-0000285.  
PR 30-JUL-2001; 2001EP-0000323.  
PR 30-JUL-2001; 2001EP-0000327.  
PR 02-AUG-2001; 2001EP-0000341.  
PR 02-AUG-2001; 2001EP-0000342.  
PR 02-AUG-2001; 2001EP-0000343.  
PR 02-AUG-2001; 2001EP-0000344.  
PR 09-AUG-2001; 2001EP-0000357.  
PR 16-AUG-2001; 2001EP-0000374.  
PR 16-AUG-2001; 2001EP-0000377.  
PR 20-SEP-2001; 2001EP-0000478.  
PR 20-SEP-2001; 2001EP-0000483.  
PR 22-OCT-2001; 2001EP-0000552.  
PR 22-OCT-2001; 2001EP-0000553.  
PR 22-OCT-2001; 2001EP-0000554.  
PR 22-OCT-2001; 2001EP-0000556.  
PR 22-OCT-2001; 2001EP-0000557.  
PR 15-NOV-2001; 2001EP-0000558.  
PR 15-NOV-2001; 2001EP-0000464.  
PR 21-DEC-2001; 2001EP-00005117.  
XX (STAM ) DSM NV.

XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;  
XX Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;  
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;  
PI Stiebler J, Albang R;  
XX WPI; 2002-723203/78.  
DR P-PSDB; ABZ78226, ABZ78283.  
XX Novel isolated protease polypeptide useful in laboratory, clinical,  
PT pharmaceutical, chemical, diagnostic, personal care and industrial  
PT applications -  
XX Claim 13; Page 366-369; 394pp; English.

XX The invention relates to a novel isolated protease polypeptide. A  
CC polypeptide or polynucleotide of the invention is useful for diagnosing a

CC fungal infection such as aspergillosis, or as a query sequence to perform  
CC a search against public databases. A polypeptide of the invention is  
CC useful in a selected number of industrial or pharmaceutical processes, in  
CC laboratory or clinical processes, in food industry (baking, brewing, in  
CC cheese manufacturing, meat tenderising), in tanning industry and in the  
CC manufacture of biological detergents. A polypeptide may also be useful  
CC for improving protein solubility, extraction yields, viscosity or taste,  
CC texture, nutritional value, minimising of antigenicity or  
CC anti-nutritional factors, colour or functionality as well as processing  
CC aspects like filterability of the proteinaceous raw material. The  
CC sequences shown in ABR38313-ABR38869 represent the A. niger proteases of  
CC the invention.

SQ	Sequence	554 AA;
----	----------	---------

Query Match 32.68: Score 941: DR 23: Length 554:

Query Match 32.6%; Score 941; DB 23; Length 554;

Best Local Similarity 38.5%; Pred. No. 9.9e-74;

Matches	209; Conservative	72; Mismatches	172; Indels	90; Gaps	13;
QY	13	OLPKNPTGVKTLTATA--NNVTIRYKEPAGAGVCVETTEGVKSYSGYVTSDESHTFFWFFE	70		
Db	17	QFPPEGITVLSKJLHENTISLFPKEP--GICETTEGVRSYSGYVHLPPASTSFFWFFE	73		
QY	71	ARHNPTAPITLMLNGPGGSDSLIGLPEELGPCHVNS--TFDDYINPHSWNEVSNLLPSQ	129		
Db	74	ARKDPSNAPLAWLNGPGGSSLMGLLEELGPCSIADSKTIVLPNWSNNEVSNLLFLDQ	133		
QY	130	PLVGFSYSYTVDGSIINPVTVGVVENSFAGVQGRYPTIDATLIDITTLNLAFAAAWEILQGF	181		
Db	134	PTOVGFSEYDVPTNGTLT-----ANCTAFAAAHALHWFQAQTW	168		
QY	190	LSGLPSL---DSRVQSKDRSLWTESYGGHYGPAFFNFHYEQNERIANGSV--NGVQ--LNFN	244		
Db	169	FEFEPHYKPNDDR-----SLMAEYSYGGHYGPGPIFRFQQQNDKIAEGTAEDGAQYLHL	223		
QY	245	SLGIINGIIDEATQAPYEPFAVNTYGIKAVNETVYNYMKFANQMENGQDILSTCKQT	304		
Db	224	TLGIIVGLMDMVLQFEAYITWP-----YNNVRLAP-----SSF	256		
QY	305	NRTALADYALCAEATNMCRD-----NV-----EGP--YYAFAGRGVVDIR	342		
Db	257	NSRGFRDQALACEAALKERDSGLPHSGKNLSEICGGLAELWGGDPTIYYITFNKGWTDIA	316		
QY	343	HPYDDTPPSYNYKFTLAKGSVMDAIGNVINYTSQNDVYVAFQQTGDFVWPNFTIEDLEE	402		
Db	317	HPKNDFFPAKHMILGYLTQESVLAALGPVNVNFTSSSSAAVATQFIKTFDVIHGGFLDA	376		
QY	403	LALPVVSLIYGADYICWFGGQAVSLAANSQAQFRSAGYTPCLKVNGVEYGETREYG	462		
Db	377	LSGVKVHVMYGDRYACNVWGEKASLAVPYSRITFEADGTYSPLLLTPDGISGMTQRLG	436		
QY	463	NFSFTRVYBAGHEVPYQPLASLQLENRRTIEGMDIAEGOKKIWPSYKNTGATATATQSS	522		
Db	437	NYSFYFVFGAGHEVFSYQVAAVEIIPRATFNKDITPGLIAYDVDEFQSVGPKDTWIKNI	496		
QY	523	VFL 525			
Db	497	PPI 499			

## RESULT 5

ABR38817

ID ABR38817 standard; Protein: 526 AA.

XX

AC	ABR
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

XX

DT 24-11-2019

XX DE

DE XX A.

KW Pro

[illegible]

(STAM ) DSM NV.

Novel isolated protease polypeptide useful in laboratory, clinical, pharmaceutical, chemical, diagnostic, personal care and industrial applications -  
P-PSDB; ABZ78184, ABZ78241.

Claim 13; Page 271-274; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A polypeptide or polynucleotide of the invention is useful for diagnosing a fungal infection such as aspergillosis, or as a query sequence to perform

CC a search against public databases. A polypeptide of the invention is  
CC useful in a selected number of industrial or pharmaceutical processes, in  
CC laboratory or clinical processes, in food industry (baking, brewing,  
CC cheese manufacturing, meat tenderising), in tanning industry and in the  
CC manufacture of biological detergents. A polypeptide may also be useful  
CC for improving protein solubility, extraction yields, viscosity or taste,  
CC texture, nutritional value, minimising of antigenicity or  
CC anti-nutritional factors, colour or functionality as well as processing  
CC aspects like filterability of the proteinaceous raw material. The  
CC sequences shown in ABR38313-ABR38869 represent the A. niger proteases of  
CC the invention.

XX SQ Sequence 526 AA;

Query Match 31.7%; Score 916.5; DB 23; Length 526;  
Best Local Similarity 38.0%; Pred. No. 1.3e-71;  
Matches 212; Conservative 78; Mismatches 187; Indels 81; Gaps 13;  
  
QY 9 VGRQLPKPTGKVLTTANNVTIRYKEPAGVCEITPGVKSYYVDTPESHTFFWF 68  
DB 16 VSRAQFVAPPDLPITKGYLDIPVKYQV-PTGICETDPSVKSFSGYVDVAEHSHTFFWF 74  
  
QY 69 FEARH-NPETAPIILWNG-----GPGSDSLIGLFEELGCPCHVNSTFDYINPHSNVEYS 122  
DB 75 FEARNDPTEAPLTVIWMGSDPGSGSSMIGLFQEHGCGCIDANGSVYNNPYSWNNAS 134  
  
QY 123 NLLFSLQPLGVGSYSDTVDSGINPVTGVVENSFAGVGQGRYPTIDAFLDITNLAAEAA 182  
DB 135 NLLYDQVQVGFYS-----IPVGYVDSSTNGEMGAFP----- 170  
  
QY 183 WEILQGLSLGLSDRSVQSKDFSLWTESYGHYGPAPFNHFEONERIANGSVNGVOLN 242  
DB 171 -----QYSRETFHTFTESYGHYGPVNEVTEQNAHLQPCA---KKIQ 211  
  
QY 243 FNSLGIINGLIIDEALQAPYPEFAV--NNTYGIKAVNETVYNYMKEANQMENGCDLIST 300  
DB 212 LGSVMIGNGWDPIIQYQAYNFTVPGNTYDYLDFNKSISLSLNNLYGPGNCJLDQYD 271  
  
QY 301 CKQNTALADYALCAEATNMCRONVGPYAFAGRGYDIDRHPYDDPTPSYNNKFLAK 360  
DB 272 C-----ARGIDEICSTADDCANEVENVYDIYSGRDEYDFRELTDFPFYFYVDYLNK 326  
  
QY 361 DSVMDAIGNNYNQSNDVYAFQOTG-DFVWPNFIEDLEILALPVRSVLSIYGDYI 419  
DB 327 ASVQAAGIAYINYESNNAVLAFSGTGDGRIMNTIQDVGKLLQGVTVVMYAGDADYN 386  
  
QY 420 CNWFGSOAVSLAANYSOAAQFSAQYTPLVK-NGVEYGETRYGNFSTRVYEAGHEVPY 478  
DB 387 CNWLGGEAVSLQV---KAANFSSAGYTNIVTSDGVTHGQVROAGQAFVRVYESGHEVPF 443  
  
QY 479 YOPTIASLQLENFTFCWDIAEGOKKIPWSYKNTGT-----ATATH-T 519  
DB 444 YOPLALEMFERVIGKOVATCKIPSSLQVGTGPKSYRBSGNTIQWEVLDSLATYNT 503  
  
QY 520 OSSVPLPATSSSSVGM 537  
DB 504 TTNAPNPVSRRLKRMGA 521

RESULT 6

ABG93049  
ID ABG93049 standard; Protein; 508 AA.

XX AC ABG93049;

DT 21-NOV-2002 (first entry)

XX DE S. cerevisiae BAX-associated protein fragment SEQ ID 56.

XX KW Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;  
KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;  
KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;  
KW neurodegeneration; cell death.

XX OS Saccharomyces cerevisiae.  
XX PN WO200264766-A2.  
XX PD 22-AUG-2002.  
XX PF 21-DEC-2001; 2001WO-EP15398.  
XX PR 22-DEC-2000; 2000EP-0870318.  
XX PR 04-JAN-2001; 2001EP-0870002.  
XX PR 09-JAN-2001; 2001EP-0870003.  
XX PA (JANC ) JANSSEN PHARM NV.  
XX PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;  
XX WP1; 2002-667002/71.  
XX DR N-PSDB; ABQ76315.  
XX CC New isolated nucleic acid representing a synthetic Bax gene, useful as  
XX PT medicament for treating, preventing and/or alleviating yeast or fungal  
XX PT infections or proliferative disorders, or for preventing apoptosis in  
XX PT certain diseases  
XX PS Claim 36; Figure 1; 344pp; English.  
XX CC This invention describes a novel nucleic acid representing a synthetic  
XX CC Bax gene. The Bax gene of the invention is useful for identifying  
XX CC Bax-resistant yeast or fungi, identifying, or obtaining and identifying  
XX CC Candida spp. sequences that are differentially expressed in a pathway  
XX CC eventually leading to programmed cell death or identifying inhibitors or  
XX CC inhibitor sequences of Bax-induced cell death. The products of the  
XX CC invention have cytostatic, fungicide, immunosuppressive, virucide and  
XX CC vasotropic activity and can be used in vaccines or for gene therapy. The  
XX CC isolated nucleic acids, polypeptides, pharmaceutical compositions,  
XX CC antisense molecules and antibodies are useful as medicaments or in  
XX CC preparing a medicament for treating, preventing and/or alleviating  
XX CC diseases associated with yeast or fungi or proliferative disorders, such  
XX CC as cancer, or for preventing apoptosis in certain diseases. The compounds  
XX CC or polypeptides, or the genetically modified organism are useful for  
XX CC preparing a medicament for modifying the endogenic flora of humans and  
XX CC other mammals. The vaccine is useful for immunising against yeast or  
XX CC fungal infections. Apoptosis-related diseases include autoimmune disease,  
XX CC ischaemia, diseases related with viral infections or neurodegenerations.  
XX CC This sequence represents a polypeptide associated with the Bax gene  
XX CC described in the disclosure of the invention.

XX SQ Sequence 508 AA;

Query Match 22.0%; Score 636; DB 23; Length 508;  
Best Local Similarity 32.6%; Pred. No. 5.6e-47;  
Matches 167; Conservative 70; Mismatches 197; Indels 78; Gaps 14;

QY 1 LPGSTPASVGRRLQ-----PKNPTGVKTLTANNVTIRYKEPAGVCEITPGVKSYSYGV 56  
DB 43 LPQNTQOTLKLDRLNHDDPLFTFTSSVDITSLRLRTVDPSKLG1----DTVKQSGYGM 98  
  
QY 57 DTSPESHTFFWFPEARHNPETAPITLWNGGPGSDSLIGLFEELGCPCHVNSTFDYINPH 116  
DB 99 DYKDSKHFFYWFESNDPANDPILWNGGPGSCSFTGLLFGPSSIGADMKPITHPY 158  
  
QY 117 SNNEVSNLLFTSOPLGVSYSYSDTVDSGINPVTGVVENSFAGVQGRYPTIDAFLDITFN 176  
DB 159 SWNNASMTIFLEQLGVGSYGD-----EKVSSTK 188  
  
QY 177 LAEAAWEILQGLSLGLSDRSVQSKDFSLWTESYGHYGPAPFNHFEYQON-ERIANGS 235  
DB 189 LACKDAYIFLELTFEAPPHL-----RSDNFIAGESYAGHYIPIQIAHEIVVKNFER----- 239  
  
QY 236 VNGVQLNFNSLGIINGIIDEALQAPYPEFAV--NNTYGIKAVNETVYNYMKEANQMENG 294  
DB 240 ----TFNLTSVMIGNGITDPLQADYIYEPMACGKGGVHPLVLSSECEKMKSAAGR----- 290

**QY** 295 QDLTSTCKQTNRITALADYL--CAEATNMCRNDNVEGYYAFAGRGVYDIRHPYDDP- - - - 340  
**Dd** 291 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
-----CRRLNKLCYASKSSLPICIVATAVCDSALLEPYIN-TGLNVYDIRGPCEDNSTDG 343  
**QY** 349 --TPPSYNNKFLLAKDSVMDAIGVINI-NYTOSNNVDVVYAFCQGTGFVWPNFIEDLEEILA 404  
**Dd** 344 | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :  
MCYTGLRYVDGYMNPPEVQETLGSVDHNYSGCDNDVFTEFTLTGDGSKP-FQCYIAELIN 402  
**QY** 405 LPVRSLIYGDAYDICNWFGGOAVSLAANYSOAAQFRSAGTIP--LKYNGVVEYGETREVG 462  
**Dd** 403 : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :  
HNIPVLIIYAGDKDVICNKWLNHAHSNELEWINKKRIYQRRLRPLRWVSKEKGDELGVKNYG 462  
**QY** 463 NFSETRVYEAGHEVPYQPTIASLOLFNRTIEG 494  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
**Dd** 463 PFTFLRIYDAGHMVPYDQPASLEWNSWSWG 494

RESULT 7  
ABR38819  
ID ABR38819 standard; Protein; 536 AA.  
XX  
AC ABR38819;  
XX  
DT 24-APR-2003 (first entry)  
XX  
DE A. niger carboxypeptidase Y polypeptide #1.  
XX  
KW Protease; fungal infection; aspergillosis; food; tanning; detergent;  
KW protein solubility; viscosity; taste; texture; nutritional value;  
KW EC3.4.16.5.  
XX  
OS Aspergillus niger.  
XX  
PN WO200268623-A2.  
XX  
PD 06-SEP-2002.  
XX  
FF 22-FEB-2002; 2002WO-EPO1984.  
XX  
PR 23-FEB-2001; 2001EP-0200657.  
PR 23-FEB-2001; 2001EP-0200658.  
PR 23-FEB-2001; 2001EP-0200660.  
PR 26-FEB-2001; 2001EP-0200706.  
PR 26-FEB-2001; 2001EP-0200707.  
PR 26-FEB-2001; 2001EP-0200708.  
PR 26-FEB-2001; 2001EP-0200719.  
PR 28-MAR-2001; 2001EP-0000075.  
PR 28-MAR-2001; 2001EP-0000078.  
PR 28-MAR-2001; 2001EP-0000080.  
PR 28-MAR-2001; 2001EP-0000087.  
PR 28-MAR-2001; 2001EP-0000088.  
PR 21-MAY-2001; 2001EP-0000156.  
PR 21-MAY-2001; 2001EP-0000159.  
PR 21-MAY-2001; 2001EP-0000160.  
PR 21-MAY-2001; 2001EP-0000162.  
PR 21-MAY-2001; 2001EP-0000165.  
PR 21-MAY-2001; 2001EP-0000166.  
PR 21-MAY-2001; 2001EP-0000168.  
PR 21-JUN-2001; 2001EP-0000240.  
PR 21-JUN-2001; 2001EP-0000242.  
PR 21-JUN-2001; 2001EP-0000244.  
PR 21-JUN-2001; 2001EP-0000246.  
PR 12-JUL-2001; 2001EP-0000280.  
PR 12-JUL-2001; 2001EP-0000285.  
PR 30-JUL-2001; 2001EP-0000323.  
PR 30-JUL-2001; 2001EP-0000327.  
PR 02-AUG-2001; 2001EP-0000341.  
PR 02-AUG-2001; 2001EP-0000342.  
PR 02-AUG-2001; 2001EP-0000343.  
PR 02-AUG-2001; 2001EP-0000344.  
PR 09-AUG-2001; 2001EP-0000357.  
PR 16-AUG-2001; 2001EP-0000374.





New Kluyveromyces yeast with modified protease gene - esp. used for high yield prodn. of recombinant protein, also DNA encoding yeast protease and derived peptide(s)

Disclousure; Page 28-31; 49pp; English.

The protease gene is to be modified in order to render it (partially) incapable of producing the natural protein; or result in a non-functional protease or in a protease with modified proteolytic activity. The modifications can be introduced in vitro or in situ by standard genetic engineering techniques or by exposure to mutagenic agents.

(Updated on 25-MAR-2003 to correct PN field.)

Query Match 19.7%; Score 568.5; DB 15; Length 491;  
Best Local Similarity 31.1%; Pred. No. 4.6e-41;  
Matches 151; Conservative 66; Mismatches 184; Indels 85; Gaps 14

QY 27 ANNVTIRKPGAGVCETTPGVKSXSGVVDTSPESTHFFWFEARRHNPTAPITLWLN 86  
DB 132 AYDLRVKTKDPSSIGI---DPAVKQKTGYLDONENKHLYFWFESRNDPVPVLWN 188  
QY 86 GPGCSLSLGLPEELGCHVNSTFDYINPHSNVEVSNLLFLSQPLGVGSYSDTVDSGN 145  
DB 189 GFGCSSLTGLFMELGPSSINKKIOPVYNDIAWNSASVIFLDQPVNVGYSNSA--- 244  
QY 146 NPVTGVENSFAVGQGRYPIDATLDTITNLAAEAAWEILQGLSLPSLDSRVQSKDF 205  
DB 245 -----VSDTVAAGROVALTLFPKQPF-----EYAKQDF 274  
QY 206 SLWTESYGGHYGAFFNFHEQEONERTANGSVNGVOLFNSLGIINGIIDAIQAIPYEPF 265  
DB 275 HIAGESYAGHYIPVFASEILSHKR-----NINLQSVLIGNLTGLTEYEYERPM 325  
QY 266 AVNTYTGKAVNETVNYMKFANPNMGCGODLISTCKQTNRALADYALCAFAFNMCRDN 325  
DB 326 ACDGDG-GYPAVLE-GSCQAMONALPR-COSMFESCYSESA-----WCVPASTYCNA 377  
QY 326 VEGPYAFAGRQYVIDRHYPDDP-----TPPSYNKFLAKDSVMDAIGNVIN-YTQSND 379  
DB 378 LLAPYQR-TGQNVDVRGKCEDSNNLCYSAMCYVS DYLNKTEIVAEVGAENVGDCNF 436  
QY 380 VYVAFQOTGDFVMP--NFIEDLEEIALPVRSVLIY-GDADYICNMFGQAVSAANYSQ 436  
DB 437 INRNFLFHGDWMKPYHRULVPGGLE-QIPV---LIYAGDAFICNWLGNKAWEALEWPG 491  
QY 437 AAQFSRAGYTPLKV-----NGVEGETREYGNFSTRVYEAGHEVPPYQPIASQLNRT 491  
DB 492 QAEYASAKLEDLVVNEHEHKKKIQGVKSHGNFTMRLYGGGHVMVMDQPESSUEFFNRW 551  
QY 492 IFG-W 495  
DB 552 LGGEW 556

RESULT 10  
AAR48059  
ID AAR48059 standard; Protein; 491 AA.  
AC AC  
XX AAR48059;  
DT  
XT  
DT 25-MAR-2003 (updated)  
DT 20-JUL-1994 (first entry)  
XX  
DE Sequence of protease C encoded by gene K1.PRC1.  
KW Protease; yeast; proteolysis.  
XX Kluyveromyces lactis.  
OS  
PN WO9400579-Al.  
XX  
XX 06-JAN-1994.  
XX  
XX 23-JUN-1993; 93WO-FR00623.  
XX  
XX 25-JUN-1992; 92FR-0007785.  
XX  
XX (RHON ) RHONE POULENC RORER SA.  
XX  
XX Fler R, Fournier A, Yeh P;  
XX WPI; 1994-026215/03.  
DR N-PSDB; AAQ55347.  
XX

New Kluyveromyces yeast with modified protease gene - esp. used for high yield prodn. of recombinant protein, also DNA encoding yeast protease and derived peptide(s)

Disclousure; Page 28-31; 49pp; English.

The protease gene is to be modified in order to render it (partially) incapable of producing the natural protein; or result in a non-functional protease or in a protease with modified proteolytic activity. The modifications can be introduced in vitro or in situ by standard genetic engineering techniques or by exposure to mutagenic agents.

(Updated on 25-MAR-2003 to correct PN field.)

Query Match 19.7%; Score 568.5; DB 15; Length 491;  
Best Local Similarity 31.1%; Pred. No. 4.6e-41;  
Matches 151; Conservative 66; Mismatches 184; Indels 85; Gaps 14

QY 27 ANNVTIRKPGAGVCETTPGVKSXSGVVDTSPESTHFFWFEARRHNPTAPITLWLN 86  
DB 62 AISLRIKPLDKSLGV---DTVQWSGYLDYQDSKHFFYWFPESENDDPVPVLWN 117  
QY 87 GPGCSLSLGLPEELGCHVNSTFDYINPHSNVEVSNLLFLSQPLGVGSYSDTVDSGN 146  
DB 118 GPGCSSFVGLFFELGPFSSIGADLKPINPYSMNSNASVIFLDQPVGVGSYG 170  
QY 147 PVTVGVENSFAVGQGRYPIDATLDTITNLAAEAAWEILQGLSLPSLDSRVQSKDFS 206  
DB 171 -----SKVTDDAAKDVIIFLDLFFERFPHL-----RNNDFH 203  
QY 207 LWTESYGGHYGAFFNFHEQEONERTANGSVNGVOLFNSLGIINGIIDAIQAIPYEPFA 266  
DB 204 ISGESYAGHYLPKIAH-----EIAVVAHDSSFLSSVLINGFTDPLTOYQYEPMA 256  
QY 267 VNNTYTGKAVNETVNYMKFANPNMGCGOD-----LISTCKQTNRALADYALCA 316  
DB 257 CGEG-GYPAVLE-----PECLOMNRNPLCLSLVDRCYKSH-----SVFSCV 298  
QY 317 EATNMCRDNVGYPYAFAGRQYVIDR---HPYDOP----IPPSYNNKFLAKDSVMDAIGV 369  
DB 299 LADRYCEQQITG-VYEKSGRNPVDIRSKCEADDGACQCBEIYISDYLNQEVQALGT 357  
QY 370 NINYTQS-NNDVYAFQQTGDFWPNFIEDLEEIALPVRSVLIYGDADYICNMFGQAV 428  
DB 358 DVSSFQCCSDVGIGFAFTGDFSP-FHOYVAELLQDDINVLITYAGDKDYICNWLGNLA 416  
QY 429 SLAANTYSOAQFRSAGYTPLKVNGBE--YGETREYGNFSTRVYEAGHEVPPYQPIASIQ 486  
DB 417 IEKLEWRYNEYKKOVLRTWKSEETDETGETEKSGPLYLYRIYDAGHWVPHDQFENS 476  
QY 487 LFNRTI 492  
DB 477 MVNSWI 482

RESULT 11  
ABG93281  
ID ABG93281 standard; Protein; 550 AA.  
XX  
XX ABG93281;  
XX  
DT 21-NOV-2002 (first entry)  
XX  
DE C. albicans BAX-associated protein fragment SEQ ID 520.  
XX  
XX Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;  
KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;  
KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;  
KW neurodegeneration; cell death.  
XX  
XX Candida albicans.

XX WO200264766-A2.  
XX 22-AUG-2002.  
XX 21-DEC-2001; 2001WO-EP15398.  
XX 22-DEC-2000; 2000EP-0870318.  
XX 04-JAN-2001; 2001EP-0870002.  
XX 09-JAN-2001; 2001EP-0870003.  
XX (JANC ) JANSSEN PHARM NV.  
XX Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;  
XX WPI; 2002-667002/71.  
XX N-ESDB; ABQ76547.  
XX New isolated nucleic acid representing a synthetic Bax gene, useful as  
XX medicament for treating, preventing and/or alleviating yeast or fungal  
XX infections or proliferative disorders, or for preventing apoptosis in  
XX certain diseases  
XX Claim 36; Figure 2; 344pp; English.  
XX This invention describes a novel nucleic acid representing a synthetic  
XX Bax gene. The Bax gene of the invention is useful for identifying  
XX Bax-resistant yeast or fungi, identifying, or obtaining and identifying  
XX Candida spp. sequences that are differentially expressed in a pathway  
XX eventually leading to programmed cell death or identifying inhibitors or  
XX inhibitor sequences of Bax-induced cell death. The products of the  
XX invention have cytostatic, fungicide, immunosuppressive, virucide and  
XX vasotropic activity and can be used in vaccines or for gene therapy. The  
XX isolated nucleic acids, polypeptides, pharmaceutical compositions,  
XX antisense molecules and antibodies are useful as medicaments or in  
XX preparing a medicament for treating, preventing and/or alleviating  
XX diseases associated with yeast or fungi or proliferative disorders, such  
XX as cancer, or for preventing apoptosis in certain diseases. The compounds  
XX or polypeptides, or the genetically modified organism are useful for  
XX preparing a medicament for modifying the endogenic flora of humans and  
XX other mammals. The vaccine is useful for immunising against yeast or  
XX fungal infections. Apoptosis-related diseases include autoimmune disease,  
XX ischaemia, diseases related with viral infections or neurodegenerations.  
XX This sequence represents a polypeptide associated with the Bax gene  
XX described in the disclosure of the invention.  
XX Sequence 550 AA;  
SQ  
Query Match 18.8%; Score 541.5; DB 23; Length 550;  
Best Local Similarity 31.3%; Pred. No. 1.3e-38;  
Matches 150; Conservative 63; Mismatches 157; Indels 109; Gaps 17;  
QY 49 VKSYGYVD-TSPESHTEFWFEARHNPTAPITLWNGGPGSDSLIGLFEELGPGCHVNS 107  
DB 133 VKQYTGILDIIDSLDKHLEFWFEPSRNDPKNDPIILWNGGPGCSSTGLFFELGPGSSINK 192  
QY 108 TDDYINSHSVENSNLLFLSOPLGWGSYSFVTDVDSINPVTVGVVENSSEFAGVQGRYPTI 167  
DB 193 TLHPYNYFISWNSASVIFLDQPVGVGYSY-----TCGDE----- 227  
QY 168 DATLIDTTNLAABAAWEILQGLSLPSLDKSLVQSKDFSLWTSYSGHYGPAFFNFYEQ 227  
DB 228 -----VKNTLTAAKDVVLELFFOKFP-----QFLTNKPHIAGESYAGHVIPAF----- 272  
QY 228 NERIANGSVQVLNENSLGII--NCIIDEATQAPYYPEFVAVNTYGIKAVNETVYNMK 285  
DB 273 -----ASEIINNADRSEFVLASVILGNIGITDPLQDSY----- 305  
QY 286 FANQMPNGC-----QDLIST--CKQTNRTALADYALCATNMCRD-----NVEGEY-- 330  
DB 306 ----KPMGCGEGYKPELVITEQDQMER----DYPRCAKLTKLCYSFQSALTCVPAQYIC 357  
QY 331 -----YAFAGRGVYDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGV-NIN-YTOS 376

DB 358 DSRLFQPYAQTGLNPYDIRKDCAEQGGNCYVEMDYLDDEYLNLDYVKEAVGASNDIDFTSC 417  
QY 377 NNDVYAFQQTGDFWENFIEDLEILALPVRVSLIYGADYICNWFEGGQAVSLAANYSQ 436  
DB 418 DDTVFRNFILGDEMKP-FQYVVAELLDNVPLIYAGDKYICNWLGNLAWVNELEYSD 476  
QY 437 AAQFRSAGYTPKLVNGVE-YGETREYGNFSTRVYEAGHEVPYIQTASLOLFNRTIFG 494  
DB 477 SEHFAPKPLQWKQDGKKAAGEVKNHKTFTLRIYDAGHMVFPDQPENALSMVNTWVQGS 535  
RESULT 12  
ABR38845  
ID ABR38845 standard; Protein: 481 AA.  
XX ABR38845;  
AC ABR38845;  
XX 24-APR-2003 (first entry)  
DE A. niger carboxypeptidase Y polypeptide #3.  
XX Protease; fungal infection; aspergillosis; food; tanning; detergent;  
XX protein solubility; viscosity; taste; texture; nutritional value;  
XX EC3.4.16.5.  
OS Aspergillus niger.  
XX WO200268623-A2.  
PD 06-SEP-2002.  
XX 22-FEB-2002; 2002WO-EP01984.  
XX 23-FEB-2001; 2001EP-0200657.  
XX 23-FEB-2001; 2001EP-0200658.  
XX 26-FEB-2001; 2001EP-0200706.  
XX 26-FEB-2001; 2001EP-0200707.  
XX 26-FEB-2001; 2001EP-0200708.  
XX 26-FEB-2001; 2001EP-0200719.  
XX 28-MAR-2001; 2001EP-0000075.  
XX 28-MAR-2001; 2001EP-0000078.  
XX 28-MAR-2001; 2001EP-0000080.  
XX 28-MAR-2001; 2001EP-0000087.  
XX 28-MAR-2001; 2001EP-0000088.  
XX 21-MAY-2001; 2001EP-0000156.  
XX 21-MAY-2001; 2001EP-0000159.  
XX 21-MAY-2001; 2001EP-0000160.  
XX 21-MAY-2001; 2001EP-0000162.  
XX 21-MAY-2001; 2001EP-0000165.  
XX 21-MAY-2001; 2001EP-0000166.  
XX 21-MAY-2001; 2001EP-0000168.  
XX 21-JUN-2001; 2001EP-0000240.  
XX 21-JUN-2001; 2001EP-0000242.  
XX 21-JUN-2001; 2001EP-0000244.  
XX 12-JUL-2001; 2001EP-0000246.  
XX 12-JUL-2001; 2001EP-0000280.  
XX 12-JUL-2001; 2001EP-0000285.  
XX 30-JUL-2001; 2001EP-0000323.  
XX 30-JUL-2001; 2001EP-0000327.  
XX 02-AUG-2001; 2001EP-0000341.  
XX 02-AUG-2001; 2001EP-0000342.  
XX 02-AUG-2001; 2001EP-0000343.  
XX 02-AUG-2001; 2001EP-0000344.  
XX 09-AUG-2001; 2001EP-0000357.  
XX 16-AUG-2001; 2001EP-0000374.  
XX 16-AUG-2001; 2001EP-0000377.  
XX 20-SEP-2001; 2001EP-0000478.  
XX 20-SEP-2001; 2001EP-0000483.  
XX 22-OCT-2001; 2001EP-0000552.  
XX 22-OCT-2001; 2001EP-0000553.  
XX 22-OCT-2001; 2001EP-0000554.



PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	26-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139753.	PR	27-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151080.
PR	21-JUN-1999;	99US-0139817.	PR	30-AUG-1999;	99US-0151303.
PR	22-JUN-1999;	99US-0139899.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140353.	PR	01-SEP-1999;	99US-0151930.
PR	23-JUN-1999;	99US-0140354.	PR	07-SEP-1999;	99US-0152363.
PR	23-JUN-1999;	99US-0140354.	PR	10-SEP-1999;	99US-0153070.
PR	24-JUN-1999;	99US-0140695.	PR	13-SEP-1999;	99US-0153758.
PR	28-JUN-1999;	99US-0140823.	PR	15-SEP-1999;	99US-0154018.
PR	29-JUN-1999;	99US-0140991.	PR	16-SEP-1999;	99US-0154039.
PR	30-JUN-1999;	99US-0141287.	PR	20-SEP-1999;	99US-0154779.
PR	01-JUL-1999;	99US-0141842.	PR	22-SEP-1999;	99US-0155139.
PR	01-JUL-1999;	99US-0142154.	PR	23-SEP-1999;	99US-0155486.
PR	02-JUL-1999;	99US-0142055.	PR	24-SEP-1999;	99US-0155659.
PR	06-JUL-1999;	99US-0142390.	PR	28-SEP-1999;	99US-0156458.
PR	08-JUL-1999;	99US-0142803.	PR	29-SEP-1999;	99US-0156596.
PR	09-JUL-1999;	99US-0142920.	PR	04-OCT-1999;	99US-0157117.
PR	12-JUL-1999;	99US-0142977.	PR	05-OCT-1999;	99US-0157755.
PR	13-JUL-1999;	99US-0143542.	PR	06-OCT-1999;	99US-0157865.
PR	14-JUL-1999;	99US-0143624.	PR	07-OCT-1999;	99US-0158029.
PR	15-JUL-1999;	99US-0144005.	PR	08-OCT-1999;	99US-0158232.
PR	16-JUL-1999;	99US-0144085.	PR	12-OCT-1999;	99US-0158369.
PR	16-JUL-1999;	99US-0144086.	PR	13-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144325.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144333.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159330.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159638.
PR	20-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0144814.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160767.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160815.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	22-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	23-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	26-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	25-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161359.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161922.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0161993.
PR	05-AUG-1999;	99US-0147192.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			

Query Match

13.3%; Score 383; DB 21; Length 482;

Best Local Similarity 26.2%; Pred. NO. 9e-25;

Matches 135; Conservative 68; Mismatches 202; Indels 110; Gaps 18;

QY

15 PKNPTGKVLTTANNVILRYKE-----PGAGVCEITPG-----VKSYSGYDVTSPESHT 64

32 PSPKRGVSSGDTSHENVIORESVFPKDKDLIQQLPGQFSDVTFKQYGYAVAVNPKAGR 91

DB

65 F--FWFFEARNHPETAPITLWLNCGPGSDSLIGLFEELGPFCHVNSTFDO--YINPHSWNEV 121

us-09-712-338-2\_copy\_19\_555.rag

92 FLYYYFVETIKPNTPLVIFWNGPGCSSLGAKFELGPFRRVHSDGKTLFRNPYSWNNE 151  
 122 SNLLFSLPGLVGFYSYSDVDGSSINVEVTGVVSSSAGVQGYRPTIDATLIDTTNLAAGA 181  
 152 ANVLFLPTPVGTGFSYS-----NSPINCQGDKAT-----AED 184  
 182 AWEILQGLSLGSLDSRVQSKDFSLWTSYSGHYGPAFFNFHYQNERIANGSVNGVOL 241  
 185 NYMFLVNLERFPPEYKGR-----DIYIAGQSYAGHYVPQLAQAILHRN-----NQTLI 232  
 242 NFNSLGIINGHIDEIAIQIAPYPEFAVNNTYGIKAVNEVYNYMKFANOMPNGCQDLISVC 301  
 233 NLRGILLGNPSLNRLEQDDEGKYKMF--SHGLIS--QQQMDNFKF-----CTD--SDL 280  
 302 KQTRTALADYALCAEATNMCRDNVEGPPYAFAGRGVYDIRHPY----- 345  
 281 YDWKCHLASOKIEAQKTHL-----DIYNIYAPLCNLSLSSEPKKCTTI 325  
 346 --DDPTPSYNYKFLAKDSVMDAIGVN-----INTQSNNDVYYAFQQTGDFVWPNFIED 398  
 326 MKADPCSGNYLKAVLNKEVQGEAHTANTKIPYEWTSNTKLLWENKDRVY--SLRPI 383  
 399 LEEILALPVRYSLYIGDADYICNWFGGQAVSLAANYSAQAFRSAGTYPLKVNQVEYGET 458  
 384 LQELMGKGVRLYNGVDLVIPFTISILAVKTMNLTVVKEWR-----PWFTGGHVGGET 438  
 459 REY--GNISFTVRYEAGHEVPPYQPTASLOLFNRRI 492  
 439 EDYKGNLTFVTVKAGHSVPTDQPIHALNIFTSEI 473

RESULT 14  
 ID AAG30064 standard; Protein: 502 AA.  
 XX  
 AC AAG30064;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 35876.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 XX EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132488.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137500.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144874.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.

Accession	Query Match	Score	DB	Length
21-OCT-1999	99US-0160814	13.38	26.28	502
22-OCT-1999	99US-0160815	26.28	26.28	502
23-OCT-1999	99US-0160980	68	110	502
24-OCT-1999	99US-0160981	68	110	502
25-OCT-1999	99US-0160989	68	110	502
26-OCT-1999	99US-0161404	68	110	502
27-OCT-1999	99US-0161405	68	110	502
28-OCT-1999	99US-0161406	68	110	502
29-OCT-1999	99US-0161359	68	110	502
30-OCT-1999	99US-0161360	68	110	502
31-OCT-1999	99US-0161361	68	110	502
32-OCT-1999	99US-0161320	68	110	502
33-OCT-1999	99US-0161992	68	110	502
34-OCT-1999	99US-0161993	68	110	502
35-OCT-1999	99US-0162142	68	110	502
36-OCT-1999	99US-0162142	68	110	502
37-OCT-1999	99US-0162142	68	110	502
38-OCT-1999	99US-0162142	68	110	502
39-OCT-1999	99US-0162142	68	110	502
40-OCT-1999	99US-0162142	68	110	502
41-OCT-1999	99US-0162142	68	110	502
42-OCT-1999	99US-0162142	68	110	502
43-OCT-1999	99US-0162142	68	110	502
44-OCT-1999	99US-0162142	68	110	502
45-OCT-1999	99US-0162142	68	110	502
46-OCT-1999	99US-0162142	68	110	502
47-OCT-1999	99US-0162142	68	110	502
48-OCT-1999	99US-0162142	68	110	502
49-OCT-1999	99US-0162142	68	110	502
50-OCT-1999	99US-0162142	68	110	502
51-OCT-1999	99US-0162142	68	110	502
52-OCT-1999	99US-0162142	68	110	502
53-OCT-1999	99US-0162142	68	110	502
54-OCT-1999	99US-0162142	68	110	502
55-OCT-1999	99US-0162142	68	110	502
56-OCT-1999	99US-0162142	68	110	502
57-OCT-1999	99US-0162142	68	110	502
58-OCT-1999	99US-0162142	68	110	502
59-OCT-1999	99US-0162142	68	110	502
60-OCT-1999	99US-0162142	68	110	502
61-OCT-1999	99US-0162142	68	110	502
62-OCT-1999	99US-0162142	68	110	502
63-OCT-1999	99US-0162142	68	110	502
64-OCT-1999	99US-0162142	68	110	502
65-OCT-1999	99US-0162142	68	110	502
66-OCT-1999	99US-0162142	68	110	502
67-OCT-1999	99US-0162142	68	110	502
68-OCT-1999	99US-0162142	68	110	502
69-OCT-1999	99US-0162142	68	110	502
70-OCT-1999	99US-0162142	68	110	502
71-OCT-1999	99US-0162142	68	110	502
72-OCT-1999	99US-0162142	68	110	502
73-OCT-1999	99US-0162142	68	110	502
74-OCT-1999	99US-0162142	68	110	502
75-OCT-1999	99US-0162142	68	110	502
76-OCT-1999	99US-0162142	68	110	502
77-OCT-1999	99US-0162142	68	110	502
78-OCT-1999	99US-0162142	68	110	502
79-OCT-1999	99US-0162142	68	110	502
80-OCT-1999	99US-0162142	68	110	502
81-OCT-1999	99US-0162142	68	110	502
82-OCT-1999	99US-0162142	68	110	502



*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: September 16, 2003, 22:54:28 : Search time 121 seconds  
(without alignments)  
1958.866 Million cell updates/sec

Title: US-09-712-338-2\_COPY\_19\_555  
Perfect score: 2887  
Sequence: 1 LPGSTPASVGRRLPKNPTG.....HTQSSVPLPTATMSVCGMA 537

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_P/US09712338/runat\_16092003\_144915\_19165/app\_query.fasta\_1.711  
-DB=Issued Patents\_NA -QPM=fastcap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09712338.scgn\_1.103 -runat\_16092003\_144915\_19165 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents\_NA :  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq : \*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq : \*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq : \*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq : \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq : \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2883	99.9	1662	3	US-08-943-714-1
2	623	21.6	2002	1	US-08-309-341-3
3	623	21.6	2002	1	US-08-608-267-3
4	623	21.6	2002	1	US-08-608-452-3
5	623	21.6	2002	1	US-08-608-224-3
6	623	21.6	2002	2	US-08-967-149-3
7	594	20.6	2068	1	US-08-309-341-1
8	594	20.6	2068	1	US-08-608-267-1
9	594	20.6	2068	1	US-08-608-452-1
10	594	20.6	2068	1	US-08-608-224-1
11	594	20.6	2068	2	US-08-967-149-1
12	568.5	19.7	2503	1	US-09-640-305-3

13	568.5	19.7	2503	1	US-08-360-673-3
14	523.5	18.1	2632	2	US-08-899-524-32
15	523.5	18.1	2632	3	US-08-329-892B-32
16	361	12.5	1551	2	US-08-828-488-4
17	361	12.5	1551	4	US-09-299-689A-4
18	348.5	12.1	1670	2	US-08-828-488-2
19	348.5	12.1	1670	4	US-09-299-689A-2
20	335	11.6	1814	4	US-09-702-705-319
21	335	11.6	1814	4	US-09-736-457-319
22	246.5	8.5	1604	1	US-08-665-966-9
23	246.5	8.5	1604	3	US-09-041-780-9
24	169.5	5.9	1263	2	US-08-828-488-6
25	169.5	5.9	1263	4	US-09-299-689A-6
26	163	5.6	1960	4	US-09-280-116-133
27	161.5	5.6	497	4	US-09-643-597-190
28	161.5	5.6	497	4	US-09-480-884A-190
29	161.5	5.6	497	4	US-09-542-615A-190
30	161.5	5.6	497	4	US-09-606-421B-190
31	154	5.3	357	4	US-09-280-116-131
32	154	5.3	607	4	US-09-280-116-130
33	130.5	4.5	522	3	US-08-998-416-160
34	127.5	4.4	3279	4	US-09-634-238-44
35	127.5	4.4	7818	4	US-09-634-238-14
36	126	4.4	1155	2	US-08-387-942C-18
37	126	4.4	12588	2	US-08-387-942C-1
38	122.5	4.2	2948	4	US-09-865-621A-6
39	122.5	4.2	3427	4	US-09-865-621A-8
40	122.5	4.2	3518	4	US-09-865-621A-5
41	122.5	4.2	3979	4	US-09-865-621A-4
42	122.5	4.2	4810	4	US-09-865-621A-3
43	122.5	4.2	5163	4	US-09-865-621A-7
44	122.5	4.2	5509	4	US-09-865-621A-1
45	122	4.2	1962	4	US-09-252-991A-4847

ALIGNMENTS

RESULT 1  
US-08-943-714-1  
; Sequence 1, Application US/08943714  
; Patent No. 6187578  
; GENERAL INFORMATION:  
; APPLICANT: Blinkovsky, Alexander  
; APPLICANT: Berka, Randy  
; APPLICANT: Rey, Michael  
; APPLICANT: Golightly, Elizabeth  
; APPLICANT: Klotz, Alan  
; APPLICANT: Mathisen, Thomas Erik  
; APPLICANT: Dammann, Claus  
; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids  
; TITLE OF INVENTION: Encoding Same  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6187578o No. 6187578disk of No. 6187578th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/943,714  
; FILING DATE: 103-OCT-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J  
; REGISTRATION NUMBER: 33,728  
; REFERENCE/DOCKET NUMBER: 4990.200-US  
; TELECOMMUNICATION INFORMATION:

102e

TELEPHONE: 212-867-0123  
 TELEFAX: 212-878-9655  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1662 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 FEATURE:  
 NAME/KEY: Coding Sequence  
 LOCATION: 1...1662  
 OTHER INFORMATION:  
 US-08-943-714-1

Alignment Scores:  
 Pred. No.: 7,61e-308 Length: 1662  
 Score: 2883.00 Matches: 536  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.86% Indels: 0  
 DB: 3 Gaps: 0

US-09-712-338-2\_COPY\_19\_555 (1-537) x US-08-943-714-1 (1-1662)

```

Qy      1  LeuProGlySerThrProAlaSerValGlyArgArgGlnLeuProLysAsnProThrGly 20
Db      55  CTTCCAGAGACTACCCGGGCTCGGTGAGAGACAGCTACCAGAACCCACCGGG 114

Qy      21  VallysThrLeuThrThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGlu 40
Db      115  GTCAGAGACTCTACACCGCAACAAATGTCACCATCCGGTACAAAGAACCCGGGCGAG 174

Qy      41  GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 60
Db      175  GGCGTCTCGGAGACTACCCGGGCTCAAAATCTACTCTGGATATGTGCGACACTCTCCC 234

Qy      61  GluSerHisThrPheThrPheGluAlaArgHisAsnProGluThrAlaProIle 80
Db      235  CAGTCCCAATCTCTCTGCTTCTCGAAGCAGACATAACCCAGAACTGCACCTATC 294

Qy      81  ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeu 100
Db      295  ACATTGTGTTGAATGGTGGCCCTGGAAGCGATTCTTTGATCGGCTCTCTCGAAGAGTTG 354

Qy      101  GlyProCysHisValAsnSerThrPheAspAspIyrIleAsnProHisSerTrpAsnGlu 120
Db      355  GGCCCTTGCCATGCAATTCGACTTTTGTATGACTACATCAACCCCTCACTCGTGGAACGAG 414

Qy      121  ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 140
Db      415  GTCICCAATTTACTATTCTCTGCCAGCATTTGGAGTCCGGCTTTTCATATAGTATAGG 474

Qy      141  ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal 160
Db      475  GTTGATGGTCCATTAAACCTGTAACITGGGGTGGTCAAAATTCGAGCTTTGACGAGGTT 534

Qy      161  GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGlu 180
Db      535  CAGGCGCGGTACCCCAACATTGATGCCACTCTGATGATACCAATCTTGGCCGAGAG 594

Qy      181  AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal 200
Db      595  GCCGCTTGGAGATCCTGCAGGATTCCTTAGTGGACTACCTAGCTTGGACTTAGGGTG 654

Qy      201  GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhe 220
Db      655  CAGCTAAGAGACTCAGTCTATGAGCGGAGAGCTATGGAGGCACTATGGTCTCTGCATTC 714

Qy      221  PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 240
Db      715  TTCATCATTTTTTACGACGAGAAATGAGAGAAATGCCAGGTAGTGTAAATGCTTCAG 774

```

```

Qy      241  LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGleAlaPro 260
Db      775  CTTAATTCAACTCTCTGGGAATTATTAAACGGCATCATCGACGAGGCGATCCAGGCCCT 834

Qy      261  TyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrVal 280
Db      835  TACTACCTGAATTCGGCTGTGAACAATACCTACCGGTATCAAGGTGTCAACGACCGCTC 894

Qy      281  TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr 300
Db      895  TACAACATCAAGAAGTTTGCACCAATGCAATGGTTGCCAGGATTTGATTCCACC 954

Qy      301  CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn 320
Db      955  TGCAACACAGACAAACCGCACCGCATTAGCTGACTACGCCCTCTCGCGCAAGCCACCAAC 1014

Qy      321  MetCysArgAspAsnValGluGlyProTyrThrAlaPheAlaGlyArgGlyValTyrAsp 340
Db      1015  ATGTGACGGGCAATGTGAGGGGCCATACACGCCCTTGTCTGGTGGTGGTGTGATGAT 1074

Qy      341  IleArgHisProTyrAspAspProThrProSerTyrTyrAsnLysPheLeuAlaLys 360
Db      1075  AATCGGCATCCATATGATGACCCGACTCCGCCAGTTATTACAAACAATTTCTGCAAG 1134

Qy      361  AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal 380
Db      1135  GACTCTGTCACTGGACGCTATCGGCTCAACATCAACTACCCACTCCCAATTAATGACGTC 1194

Qy      381  TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu 400
Db      1195  TACTACGCTTCCAGCAACAGGCGACTTGTCTGGCCCACTTCTCATCGAGACCTCGAG 1254

Qy      401  GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys 420
Db      1255  GAGATCTTGTCTCCCGTGGTGTCTCCCTCATCTATGGCGACGCGGATTTACATCTGC 1314

Qy      421  AsnTrpPheGlyGlyGluAlaValSerLeuAlaAlaAsnTyrSerGlnAlaGlnPhe 440
Db      1315  AACTGGTTCGGCGGTGAGCGGCTTTCCTCTCGTGGGAATCTCTCCCAAGCCCGCATTC 1374

Qy      441  ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu 460
Db      1375  CGAAGCGGAGGTGACAGCCCTGAAGTCAACGGCGTGGAGTATGGGAAACTCCGAG 1434

Qy      461  TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTrpGln 480
Db      1435  TATGTAATTTCTCTTCACTCGCGCTATAGAGCGAGGCCATGAAGTCCCATACACAG 1494

Qy      481  ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGly 500
Db      1495  CCCATCGCTCCCTGCAATTTTAACCGGACTATCTTCGGTGGGATATCGCAGAGGC 1554

Qy      501  GlnLysLysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGln 520
Db      1555  CAGAAGAAGATCTGCCCGCAGCTACAGAGATGGAACGGCTACACCTACGACATACAG 1614

Qy      521  SerSerValProLeuProThrAlaThrSerMetSerSerValGlyMet 536
Db      1615  TCGTCCGTCCCGTCCCTACGCTACCAAGCATGTCCAGTGTGGTATG 1662

```

# RESULT 2

US-08-309-341-3  
 ; Sequence 3, Application US/08309341  
 ; Patent No. 5594119  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yaver, Debbie Sue  
 ; APPLICANT: Thompson, Steryl Ann  
 ; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 5594119 No. 5594119disk of No. 5594119th America, Inc.  
 ; STREET: 405 Lexington Avenue, Suite 6400

CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10174-6401  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
 CURRENT APPLICATION DATA: US/08/309,341  
 APPLICATION NUMBER: US/08/309,341  
 FILING DATE: 16-SEP-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lowney, Karen A.  
 REGISTRATION NUMBER: 31,274  
 REFERENCE/DOCKET NUMBER: 4247.000-US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212 867 0123  
 TELEFAX: 212 867 0298  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2002 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 ORIGINAL SOURCE:  
 ORGANISM: Aspergillus niger  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 349..411  
 NAME/KEY: CDS  
 LOCATION: join (348..412)  
 US-08-309-341-3

## Alignment Scores:

Pred. No.: 1,29e-58 Length: 2002  
 Score: 623.00 Matches: 155  
 Percent Similarity: 48.25% Conservative: 79  
 Best Local Similarity: 31.96% Mismatches: 175  
 Query Match: 21.58% Indels: 76  
 DB: 1 Gaps: 18

US-09-712-338-2\_COPY\_19\_555 (1-537) x US-08-309-341-3 (1-2002)

QY 27 AlaAsnValThrThrLeArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46  
 DB 533 GCCTATGATCTCAGGTCAGAGACCGATCTCTTGGCATC-----GAC 583  
 QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerPro---GluSerHisThrPhe 65  
 DB 584 CTGGCGCTTAAGCAGTACACCGGTTATCTCGATGACACGAGAACGACGACGATCTGTC 643  
 QY 66 PheTrpPhePheGluAlaArgHisAsuProGluThrAlaProIleThrLeuTrpLeuAsn 85  
 DB 644 TACTGGTTCTTCAGTCTCGCATGACCCCGAGAAATGACCCCTGTTGTCGTGGCTGAAC 703  
 QY 86 GlyGlyProGlySerAspSerLeuLeuGlyLeuPheGluGluLeuGlyProCysHisVal 105  
 DB 704 GTGGCCCTGGATGCTCTTCCCTACCGGTCCTTTTCATGAGCTGCGCCCTACGACATC 763  
 QY 106 AsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeu 125  
 DB 764 ACAAGAGATCAGCCGGTCTACACGACTACGCTTGAACCTCCACGCGTCCGTCATC 823  
 QY 126 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 145  
 DB 824 TTCCTGACCGCTGTCAGGTCGGTTACTCTTCACGCACTCTCCT----- 871  
 QY 146 AsnProValThrGlyValValGluAlaAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 165

DB 871 ----- 871  
 QY 166 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaThrProGluIle 185  
 DB 872 -----GICAGGACACCGCTGCTGCGCAAGGAGCTATGCCCTTG 913  
 QY 186 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 205  
 DB 914 CTTACCTCTTCTTCAACAATTCCTC-----GAGTATGCCAAGCAGGACTTC 961  
 QY 206 SerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyr 225  
 DB 962 CACATGCCGGTGAATCTATGCTGGTCACTATATCCCGCTCTTGTCTCGGAGATTTC 1021  
 QY 226 GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 245  
 DB 1022 TCTCACAAGAGCGC-----AACATCAACCTGCAGTCC 1054  
 QY 246 LeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPhe 265  
 DB 1055 GTCCTTAITGGCAACGGTCTCACCGACGGTCTCACTAGTACGAGTACTACCGTCCCATG 1114  
 QY 266 AlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLys 285  
 DB 1115 GCCTGTGTGACGGT---GGTACCAGCTGCTTGGACGAG---GGCTCTGCCAGGCC 1168  
 QY 286 PheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsn 305  
 DB 1169 ATGGACAACAGCCCTTCTCTCGC---TGGCAGTCTATGATTAGTCTTCTATAGTTCGCG 1225  
 QY 306 ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsn 325  
 DB 1226 AGCGCT-----TGGGTTGTGTCGCCGCCCTCACTACTACTATAACAGGCC 1270  
 QY 326 ValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyr 345  
 DB 1271 CTCCTTCCCTTACCAGCGC---ACCGACAGACGCTACGATGCTTCGTGGTAAGTGC 1327  
 QY 346 AspAspPro-----ThrProSerTyrTyrAsnLysPheLeuAlaLys 360  
 DB 1328 GAGGATAGCTCCAACTCTGCTACTGCGCATGGCTACGTCAGTCAGCGACTACTCTGAACAAG 1387  
 QY 361 AspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsp 379  
 DB 1388 ACCGAGTCAATGAGGCTGTGGCGCTGAGGTCACGCGCTACGACTCGTCAACTTTCAC 1447  
 QY 380 ValTyrTyrAlaPheGlnGlnThrGlyAspPheValTrpPro-----AsnPheIleGlu 397  
 DB 1448 ATCAACGCGAACITCCTCTTCCACGGTGTGCTGGATGAGCGCTACCACTGCTGCTTCGG 1507  
 QY 398 AspLeuGluGluLeuLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAla 416  
 DB 1508 GGACTCTCTGGAG-----CAGATCCCTGTCTC-----CTGATCTACGCTGGTACGCC 1552  
 QY 417 AspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGln 436  
 DB 1553 GATTTATCTGCACTGGCTGGCGCAACAGCGCTGGAGCCCTTGATGAGTGGCCCGGA 1612  
 QY 437 AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn 451  
 DB 1613 CAGGCTCAATATGCTTCCCTTAAGCTGGAGGACCTGCTCGTGGTCGAGAAATGACACAC 1672  
 QY 452 GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu 471  
 DB 1673 GCGAAGAAGATCGCCAGGTCAGTCCCATGGCACTTCCACTTCACTGCTCTCTATGGC 1732  
 QY 472 AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr 491  
 DB 1733 GTGGGCCCATGCTCCGATGGACCAACCGAGTCTGAGTCTTGAAATCTTCAACCGCTGG 1792  
 QY 492 IlePheGly---Trp 495  
 DB 1793 TTGGGAGGTGAATGG 1807

## RESULT 3

```

US-08-608-267-3
; Sequence 3, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5688663o No. 5688663disk of No. 5688663th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,267
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 349..411
; NAME/KEY: CDS
; LOCATION: join (348..412)
US-08-608-267-3

Alignment Scores:
Pred. No.: 1.29e-58 Length: 2002
Score: 623.00 Matches: 155
Percent Similarity: 48.25% Conservative: 79
Best Local Similarity: 31.96% Mismatches: 175
Query Match: 21.58% Indels: 76
DB: 1 Gaps: 18

US-09-712-338-2_COPY_19_555 (1-537) x US-08-608-267-3 (1-2002)
Qy 27 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46
Db 533 GCCTATGATCTCAGCGTCAGAGACCGATCCAGTCTCTTGGCATC-----GAC 583
Qy 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerPro----GluSerHisThrPhe 65
Db 584 CCTGGCGTAAAGCAGTACACCGGTATCTCGATGACACGAGACGACGACGATCTGTTC 643
Qy 66 PheIrrPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsn 85
Db 643 PheIrrPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsn 85

```

```

Db 644 TACTGGTCTTCGAGICTCCGAATGACCCCGAGAAIGACCCCTGTGTCTGTGGCTGAAC 703
Qy 86 GlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisVal 105
Db 704 GGTGGCCCTGATGCTCTTCCCTCAGCGGTCTTTTCATGAGCTGCGCCCTAGCAGCATC 763
Qy 106 AsnSerThrPheAspAspTyrIleAsnProHisSerTirpAsnGluValSerAsnLeuLeu 125
Db 764 AACAAAGAATCCAGCGGTCTACACGACTACGCTTGGAACTCCAAAGCGTCCGTGATC 823
Qy 126 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 145
Db 824 TTCCTTTGACGAGCTGTCAACGCTGGTGTACTTACAGCACTCTGT----- 871
Qy 146 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgIyrPro 165
Db 871 ----- 871
Qy 166 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaTrpGluIle 185
Db 872 -----GTCAGCGACACGGTGTCTGCTGCAAGAGAGCTCTATGCCCTTG 913
Qy 186 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 205
Db 914 CTTACCCCTCTCTTCAACAATTCCTC-----GAGTATGCCAAGCAGCAGCTTC 961
Qy 206 SerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsnHisPheTyr 225
Db 962 CACATTGCGGTGAATCTTATGCTGTCACATATATCCCGTCTTTGCTTGGAGATTTG 1021
Qy 226 GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 245
Db 1022 TCTCACAAGAAGCGC-----AACATCAACCTGCATCC 1054
Qy 246 LeuclyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPhe 265
Db 1055 GTTCTTATTTGCAACGGTCTCACCGAGCGTCTACTCAGTACGAGTACTACCGTCCCATG 1114
Qy 266 AlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetlys 285
Db 1115 GCCTGTGTGACGGT---GGTTACCCAGCTCTCTGGACGAG---GGTCTCTGCCAGGCC 1168
Qy 286 PheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsn 305
Db 1169 ATGGACAACGCCCTCTCTCGC---TGCAGTCTATGATGAGTGTCTGTATAGTTCGAG 1225
Qy 306 ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsn 325
Db 1226 AGCGCT-----TGGTGTGTGTCCCGGCTCCATCTACTCTACTTAACAACGCC 1270
Qy 326 ValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyr 345
Db 1271 CTCCTTGGCCCTTACCAGCGC---ACCGGACAGAACGCTACGATGTTCGTGTAAGTGC 1327
Qy 346 AspAspPro-----ThrProSerTyrTyrAsnLysPheLeuAlaLys 360
Db 1328 GAGATAGCTCCAAACCTCTGTCTACTCGGCCGAGGGCTACGTCAGCAGCTACTCTCAACAG 1387
Qy 361 AspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsuAsnAsp 379
Db 1388 ACCGAGGTCAATTGAGGCTGTGGCTGAGTCAACGGCTACGACTCGTCAACTTGGAC 1447
Qy 380 ValTyrTyrAlaPheGlnGlnThrGlyAspPheValTrpPro-----AsnPhelleGlu 397
Db 1448 ATCAACGCCAACTTCTCTTCCAGGTGACTGGATGAAGCCCTACCCGCTCTGTTCCG 1507
Qy 398 AspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAla 416
Db 1508 GGACTCTCGGAG-----CAGATCCCTCTC-----CTGATCTAGCTGGTGACGCC 1552
Qy 417 AspTyrIleCysAsnThrPheGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGln 436
Db 1553 GATTTCACTGCACCTGCTGGCACAACAGGCTTGACTGAAGCCCTTGAGTGGCCCGGA 1612

```



```
QY 361 AspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsnAsp 379
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1388 ACCGAGTATGAGGCTGTGGCGTGAAGTCAACGGCTACGACTCGTCCACTTTGAC 1447
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 380 ValTyrAlaPheGlnGlnThrGlyAspPheValTrpPro-----AsnPheIleGlu 397
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1448 ATCAACCGCAACTTCTCTTCCACGGTGACTGGATGAAGCCCTACCACCGTCTCGTTCCG 1507
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 398 AspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAla 416
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1508 GGACTCTCTGGAG-----CAGATCCCTGTC-----CTGATCTACGCTGTGACGCC 1552
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 417 AspTyrIleCysAsnTrpPheGlyGlnAlaValSerLeuAlaAsnTyrSerGln 436
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1553 GATTTCATCTGCNACTGGCTGGGCAACAGGCTTGACTCAAGCCCTTGAGTGCCCGGA 1612
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 437 AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn 451
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1613 CAGGCTGAATATGCCCTCCGCTAAGCTGGAGGACTGGTGGTGGAGAAATGAGCACAAG 1672
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 452 GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu 471
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1673 GGCAAGAAGATCGCGCAGGTCAAGTCCCATGGCAACTTCACCTTCATGGCTCTCTAIGGC 1732
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 472 AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr 491
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1733 GGTGGCCACATGTCTCCCGATGGACCAACCCGAGTCGAGTCTTGAAATCTTCAACCGCTGG 1792
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 492 IlePheGly---Trp 495
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1793 TTGGGAGGTGAATGG 1807
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 5
US-08-608-224-3
; Sequence 3, Application US/08608224
; Patent No. 5705376
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 57053760 No. 5705376disk of No. 5705376th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,224
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 349..411
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (348..412)
US-08-608-224-3

Alignment Scores:
Pred. No.: 1,29e-58 Length: 2002
Score: 623.00 Matches: 155
Percent Similarity: 48.25% Conservative: 79
Best Local Similarity: 31.96% Mismatches: 175
Query Match: 21.58% Indels: 76
DB: 1 Gaps: 18

US-09-712-338-2_COPY_19_555 (1-537) x US-08-608-224-3 (1-2002)
QY 27 AlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 533 GCCTATGATTCAGGCTCAAGAGACGATCCAGCTCTCTTGGCATC-----GAC 583
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerPro---GluSerHisThrPhe 65
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 584 CTGGCGTAAAGCAGTACACCGGTTAICTGATGACACGAGACGACGACATCTGCTTC 643
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 66 PheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsn 85
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 644 TACTGTCTTCGAGTCTCGCAATGACCCGAGAAATGACCTGTGTCTGTGGCTGAAC 703
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 86 GlyClyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisVal 105
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 704 GGTGGCCCTGGATGCTCTCCCTCACCCTGCTTTTCATGGAGCTCGGCCCTAGCAGCATC 763
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 106 AsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeu 125
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 764 AACAGAAGATCCAGCCGGCTACACAGCACTACGCTTGAACCTCAACGGCTCGTGATC 823
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 126 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 145
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 824 TTCTTACACGAGCTGTCACACGCTGGTTACTCTTACAGCACTCTGCT----- 871
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 146 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 165
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 871 ----- 871
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 166 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaIleTrpGluIle 185
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 872 -----GTCAGGACACCGCTGCTGCTGCAAGGAGCTGCTATGCTTG 913
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 186 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 205
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 914 CTTACCTCTCTTCAAAACAATTCCCC-----GAGTATGCCAAGCAGGACTTC 961
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 206 SerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsnHisPheTyr 225
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 962 CACATTGCCGGTGAATCCTATGCTGGTCACTATATCCCGCTCTTGGCTCGGAGATTGG 1021
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 226 GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 245
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1022 TCTCACAGAAGCGC-----AACATCAACCTGCAGCTCC 1054
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 246 LeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTrpProGluPhe 265
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1055 GTTCTTATGGCAACGCTCTACCGAGGCTCTACTCAGTACGAGTACTACCGTCCCATG 1114
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 266 AlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLys 285
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1115 GCCTGTGTGACGGT---GGTTACCCAGCTGCTTGGACGAG---GGCTCTCTGCCAGGCC 1168
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

```

QY      286 PheAlaasnGlnMetProasnGlyCysGlnAspLeuileSerThrCysLysGlnThrAsn 305
Db      1169 ATGGACAAGCCCTCTCTCGC---TCCAGTGTATGATGAGTCTTGTCTATATGTTCCGAG 1225
QY      306 ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsn 325
Db      1226 AGCGCT-----TGGGTTTGTGTCGCCGCTCCATCTACGTACGTAAACAACGCC 1270
QY      326 ValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyr 345
Db      1271 CTTCTTGCCCTTACACAGCC---ACCGAGACAGAGCTGTACGATGTGCTGTAAGTGC 1327
QY      346 AspAspPro-----ThrProSerTyrTyrAsnLysPheLeuAlaLys 360
Db      1328 GAGGATAGTCCAACTCTGCTACCTGGCAGTGGCTAGTACGACACTACCTACCAACAG 1387
QY      361 AspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsp 379
Db      1388 ACCGAGGTGATGAGGCTGTGGCGCTGAGGTCAACGGCTACGACTCGTGCAACTTIGAC 1447
QY      380 ValTyrTyrAlaPheGlnGlnThrGlyAspPheValTrpPro-----AsnPheIleGlu 397
Db      1448 ATCAACCGCACTCTCTTCACCGGTGACTGGATGAAGCCCTACCAACCGTCTCGTTCCG 1507
QY      398 AspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAla 416
Db      1508 GGACTCTCGAG---CAGATCCTGTC-----CTGATCTACCGTGTGAGCGC 1552
QY      417 AspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGln 436
Db      1553 GATTCACTGCACTGGCTGGCAACAAAGGCTGGACTGAAGCCCTTGAGTGGCCCGCA 1612
QY      437 AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn 451
Db      1613 CAGGTGATGATGCTCCGCTGAAGCTGGAGGACCTGGTGTGTCGAGAAATGAGCACAA 1672
QY      452 GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu 471
Db      1673 GGCAAGAAGATCGCCGAGTCCGATGCCATGCCATGCCACTTCACCTTCATGCTCTAAGC 1732
QY      472 AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr 491
Db      1733 GGTGGCCCATGCTCCGATGACCAACCCGAGTGGAGTCTTGAATCTTCAACCGCTGG 1792
QY      492 IlePheGly---Trp 495
Db      1793 TTGGAGGTGAATGG 1807

```

## RESULT 6

```

US-08-967-149-3
; Sequence 3, Application US/08967149
; Patent No. 5939305
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5939305 No. 5939305disk of No. 5939305th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,149

```

```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,452
; FILING DATE: 28-FEB-1996
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 349..411
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (348..412)
; US-08-967-149-3

```

```

Alignment Scores:
Pred. No.: 1,29e-58 Length: 2002
Score: 623.00 Matches: 155
Percent Similarity: 48.25% Conservative: 79
Best Local Similarity: 31.96% Mismatches: 175
Query Match: 21.58% Indels: 76
DB: 2 Gaps: 18

```

```

US-09-712-338-2_COPY_19_555 (1-537) x US-08-967-149-3 (1-2002)

QY      27 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46
Db      533 GCCTATGATCTCAGGCTCAAGAACCGATCTCTCTTGTGGCATC-----GAC 583
QY      47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerPro---GluSerHisThrPhe 65
Db      584 CCTGGCGTAAAGCAGTACACCGGTTATCTCGATGACACGAGAACGACGACATCTGTT 643
QY      66 PheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsn 85
Db      644 TACTGGTCTTCGAGTCTCGCAATGACCCCGAGATGACCCCTGTTCTTGTGGCTGAAC 703
QY      86 GlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisVal 105
Db      704 GTGGCCCTGGATGCTCTTCCCTCACCAGTCTTTCATGAGCTCGGCCCTAGCACATC 763
QY      106 AsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu 125
Db      764 AACAAGAAGATCCAGCGGCTCTACACGACTACGCTTGGAACTCCCAACGCGTCCGTCATC 823
QY      126 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 145
Db      824 TTCCTTGACAGCCCTGCAACGTCGGTTACTTACTTACAGCACTCTGCT----- 871
QY      146 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 165
Db      871 ----- 871
QY      166 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIle 185
Db      872 -----GTCAGGACACCGCTGTGCTGTCGTCGACGAGGACGCTCTAIGCCCTG 913

```





QY		105	ValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu	124
Dd		747	ATCAACAAGAAGATCCACCGGTCTACAATAACTACGTTGGAACTCCAAGCGTCGGTG	806
QY		125	LysPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySer	144
Dd		807	ATCTTCCTCAGCACGCCTGTCATGTGCGGTACTCTACAGTAAGTCTGCT---	857
QY		145	IleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyr	164
Dd		857	-----	857
QY		165	ProThrIleAspAlaThrLeuIleAspThrThrAsnLeualaLaGluAlaAlatrPrGlu	184
Dd		858	-----GTACAGCAGACCGGTGCGTCTGCTGCCAAGCAGCGTCTATGCC	896
QY		185	IleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAsp	204
Dd		897	TTCGTTACCCTCTCTCAAACAATTCCC- - - - - GAGTAGTGIAAGCAGGAC	944
QY		205	PheSerLeuTrpThrGluSerTyrGlyGlyHistyGlyProkAlaphePheAsnHisPhe	224
Dd		945	TTCACCATTCGCCGTGAATCTTATGTGTGCTACTATATCCCGTCTTCGTTCCGAGATC	1004
QY		225	TyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsn	244
Dd		1005	CTGTCTCACAGAAGCGC- - - - - AACATCAACCTGCAG	1037
QY		245	SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIlegInAlaProTyrTyrrProGlu	264
Dd		1038	TCCGTTCTCATTTGGCAACGGTCTCACCGCGGATACACCCAGTACGAGIACACCGTCC	1097
QY		265	PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMet	284
Dd		1098	ATGCGTGGCGGTACGCGCGGTATCCACGCTGTCTTGACGAGAGCTCTGCCAGTCCAT	1157
QY		285	LysPheAlaAsnGlnMetProAsnGlyCysGlnAspleuIleSerThrCysLysGlnThr	304
Dd		1158	GAC- - - - - AAGCTCTTCTCCG- - - TGCAGTCTATGATGAGICTTGCTACAGTTC	1208
QY		305	AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGlualathrAsnMetCysArgAsp	324
Dd		1209	GAGAGCGCT- - - - - TGSGTTGTGTCCTCCGSCCTCCATCTACTGTACAAC	1253
QY		325	AsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspileArgHisPro	344
Dd		1254	GCCCTCTGCCCTTACCAGCGC- - - ACTGGCAGAACGCTATATGATGCTCGGTGAAG	1310
QY		345	TyrAspAspPro- - - - - ThrProProSerTyrTyrAsnLysPheLeuAla	359
Dd		1311	TGGAGGATAGCPCTAACCTTTGCTACTCGGCTATGGCTACGICAGGNACTACCTGAAC	1370
QY		360	LysAspSerValMetAspAlalleGlyValAsnIleAsn- - - TyrThrGlnSerAsnAsn	378
Dd		1371	AAGCCGAACTCATCGAGGCTGTGGCGGTGAGGTCAACGGCTACGACTCGTCCAACCTT	1430
QY		379	AspValTyrTyrAlaPheGlnThrGlyAspPheValTrppro- - - - - AsnPhelle	396
Dd		1431	GACATCAACCGCNACTCTCTTCCACGGGTACTGTGATGAAGCCCTACCAACGCCCTCGT	1490
QY		397	GluAspLeuGluGluileLeuAlaLeuProValArgValserLeulleTyr---GlyAsp	415
Dd		1491	CCGGGACTCTGGAG- - - - - CAGATCCCTGTGTC- - - - - TTGATCTATCGCGTGAT	1535
QY		416	AlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValserLeuAlaAlaAsnTyrSer	435
Dd		1536	GCTGATTTGATTTGCAACTGCTGGGCAACAAGCGCTGAGTCAAGGCCCTGGAGIGGCC	1595
QY		436	GlnAlaAlacIlnPheArgSerAlaGlyTyrThrProLeuLysVal- - - - -	450
Dd		1596	GGACAGGCTGAATATGCCCTGCTAGCTGGAGGATCTGCTATCTTGCNACAATGAGCAC	1655

```

QY 451 AsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyr 470
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1656 ACGGCAAGAGATGTGCCAGGTTAAGTCCCATGGCAACITTCACCTTCATGCGTCTCTAT 1715
QY 471 GluAlaGlyHisGluValProTyrTyrGlyProPheLeuAlaSerLeuGlnLeuPheAsnArg 490
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1716 GGTGGTGGCCACATGGTCCGATGACGACCGAGTGCAGTCTCGAGTCTTTCACCGC 1775
QY 491 ThrIlePheGly---Trp 495
||| ||| ||| |||
Db 1776 TGGTGGGAGTGAATGG 1793

RESULT 8
US-08-608-267-1
; Sequence 1, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 56886630 No. 5688663disk of No. 5688663th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,267
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 572..632
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join (571..633)
; US-08-608-267-1

Alignment Scores:
Pred. No.: 2.14e-55 Length: 2068
Score: 594.00 Matches: 155
Percent Similarity: 46.44% Conservative: 80
Best Local Similarity: 30.63% Mismatches: 174
Query Match: 20.57% Indels: 98
DB: 1 Gaps: 18
US-09-712-338-2_COPY19_555 (1-537) x US-08-608-267-1 (1-2068)

```

```
QY 27 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 GCCTATGATCTCAGGCTCAAGAGCAGGATCGCTGCTCTTGCGATC-----GAC 507
QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGlu----- 61
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 508 CCGCGGCTGAAGCAGTACACCGGTTATCTCGATGACACGA-CAATGATAAGCATTTGTT 566
QY 62 -----SerHisThrPhePhe----- 66
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 567 CTAAGTAACACACCTTGTTCAAGATCAGCGCTTTTATATGCTCTGGATATCTAACGCA 626
QY 67 -----TrpPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeu 84
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 627 ACTTAGTGTTCTTCAGTCTCGCATGACCCGAGAAATGATCCCGTTGTTCTGTTGCTG 686
QY 85 AsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHis 104
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 687 AACGGTGGCCCTGGGTGCTCTCCCTCAGCGGTCTCTTCATGGAGCTTGCCCTAGCAGC 746
QY 105 ValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu 124
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 747 ATCAACAAGAGATCCAGCGGTCTACAATGACTACGCTTGGAACTCCAACGGCTCCGG 806
QY 125 LeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySer 144
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 807 ATCTTCCTTGACCGAGCTGCAATGTCGGTTACTCTCTACAGTAACCTCTGCT----- 857
QY 145 IleAsnProValThrGlyValValGluValGluAsnSerSerPheAlaGlyValGlnGlyArgTyr 164
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 857 ----- 857
QY 165 ProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaTrpGlu 184
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 858 -----GTCAGCAGACAGCTGCTGCTGCGAAGGACGCTATGCGC 896
QY 185 IleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAsp 204
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 897 TTGCTTACCTCTCTTCAACAATTCCTC-----CAGTATGCTAAGCAGGAC 944
QY 205 PheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsnHisPhe 224
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 945 TTCACATTGCCCGTGAATCTTATGCTGTGCTACATATCCCTCTCGCTCGGAGATC 1004
QY 225 TyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsn 244
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1005 CTGCTCAACAAGAGCGC-----ACATCAACCTGCGAG 1037
QY 245 SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTrpGlu 264
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1038 TCGGTTCTCATTTGGCAAGGCTCTCAGGAGGATACACCCAGTACGAGTACTACCGTCC 1097
QY 265 PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMet 284
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1098 ATGCGCTCGGCTGAGCGGTTACCCAGCTGTCTGGACGAGAGCTCTGCCAGTCCATG 1157
QY 285 LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr 304
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1158 GAC-----AACGCTCTTCCTCGG--TGCCAGTCTATGATTCAGTCTTCTACAGTCC 1208
QY 305 AsnArgThrAlaLeuAlaAspTyrAlaLysCysAlaGluAlaThrAsnMetCysArgAsp 324
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1209 GAGAGCGCT-----TGGTGTGTTGTCGCCGCCCTCCACTACTCTGTAACAAC 1253
QY 325 AsnValGluGlyProTyrThrAlaPheAlaGlyValTyrAspIleArgHisPro 344
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1254 GCCTCTCTCCCTTACCAGCGC---ACTGGCAGAACGCTATGATGTCGCTGGTAAG 1310
QY 345 TyrAspAspPro-----ThrProSerTyrThrAsnLysPheLeuAla 359
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1311 TGCAGGATAGCTTACCTTTGCTACTCGGCTATGGGCTACGTCAGCGACTACTCTGAC 1370
```

## RESULTS 9

US-08-608-452-1  
; Sequence 1, Application US/08608452  
; Patent No. 5693510

## GENERAL INFORMATION:

; APPLICANT: Yaver, Debbie Sue  
; APPLICANT: Thompson, Sheryl Ann  
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 56935100 No. 5693510disk of No. 5693510th America, Inc.  
; STREET: 405 Lexington Avenue, Suite 6400  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA: US/08/608,452  
; APPLICATION NUMBER: US/08/608,452  
; FILING DATE: 28-FEB-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/309,341  
; FILING DATE: 20-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lowney, Karen A.  
; REGISTRATION NUMBER: 31,274  
; REFERENCE/DOCKET NUMBER: 4247.000-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 867 0123  
; TELEFAX: 212 867 0298  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2068 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA  
 ORIGINAL SOURCE: Aspergillus niger  
 FEATURE: intron  
 NAME/KEY: intron  
 LOCATION: 572..632  
 FEATURE: CDS  
 NAME/KEY: join (571..633)  
 LOCATION: 571..633  
 US-08-608-452-1

## Alignment Scores:

Pred. No.: 2,14e-55 Length: 2068  
 Score: 594.00 Matches: 155  
 Percent Similarity: 46.44% Conservative: 80  
 Best Local Similarity: 30.63% Mismatches: 174  
 Query Match: 20.57% Indels: 98  
 DB: 1 Gaps: 18

US-09-712-338-2\_COPY\_19\_555 (1-537) x US-08-608-452-1 (1-2068)

QY 27 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46  
 DB 457 GCCTATGATCTCAGGGTCAAGAGACCGATCTGGCTCTCTTGGCATC-----GAC 507  
 QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGlu----- 61  
 DB 508 CCGGGCGTGAAGCAGTACACGGTTATCTCGATGACACAGCA-GAATGATAGCAATTGTT 566  
 QY 62 -----SerHisThrPhePhe----- 66  
 DB 567 CTACGTAAAGCACACCTGGTTCAGATCACCGCTTTTATATGCTCTGGATCTAACGCA 626  
 QY 67 -----TrpPheGluAlaArgHisAsnProGluThrAlaProIleIhrLeuTrpLeu 84  
 DB 627 ACTTAGTGTCTCGAGTCTCGAATGACCCGAGAGATCATCCGTTGCTTGTGTGCTG 686  
 QY 85 AsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHis 104  
 DB 687 AACGGTGGCCCTGGGTCTCTCCCTACCGGCTCTCTTCATGAGCTTGGCCCTAGCAGC 746  
 QY 105 ValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu 124  
 DB 747 ATCAACAAGAGATCCACCGCGGTCTACAATGACTACGCTTGAACCTCAACGCGTCCG 806  
 QY 125 LeuPheLeuSerGlnProGluValGlyPheSerTyrSerAspThrValAspGlySer 144  
 DB 807 ATCTTCTTACGACCGCTGTCATGTGGTTACTCTTACAGTAACTCTGCT----- 857  
 QY 145 IleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyr 164  
 DB 857 ----- 857  
 QY 165 ProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaIleTrpGlu 184  
 DB 858 -----GTCACGACACCGTCTGCTGCTGGCAGGAGCTATGCC 896  
 QY 185 IleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAsp 204  
 DB 897 TTGCTTACCTCTCTTCAACAAATTCCTCC-----GAGTATGCTAAGCAGGAC 944  
 QY 205 PheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsnHisPhe 224  
 DB 945 TTCACATTGCGCGTGAATCTTATGTGTCATCTATATCCCGCTCTCGGATC 1004  
 QY 225 TyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsn 244  
 DB 1005 CTGCTCTCAAGAAGAGCG-----ACATCAACCTGCAG 1037  
 QY 245 SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrProGlu 264  
 DB 1038 TCGGTTCTATTGGCAACGCTCTCACCGGGAATACACCCAGTACGAGTACTACCGTCC 1097

QY 265 PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMet 284  
 DB 1098 ATGCCICGGGTGAGCGGGTTACCCAGCTGTCTGGAGAGAGCTCTGCCAGTCCATG 1157  
 QY 285 LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr 304  
 DB 1158 GAC-----AACGCTCTTCCCTCCG---TGCCAGCTATGATGAGTCTTGTACAGTTC 1208  
 QY 305 AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaIhrAsnMetCysArgAsp 324  
 DB 1209 GAGAGCGCT-----TGGGTGTGTCCCGGCTCCATCTACTGTAAACAC 1253  
 QY 325 AsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisPro 344  
 DB 1254 GCCCTCTTGGCCCTTACACGCGC---ACTGGCAGAACGCTATGATGCTCGGTGAAG 1310  
 QY 345 TyrAspAspPro-----ThrProSerTyrTyrAspLysPheLeuAla 359  
 DB 1311 GCGAGGATAGCTCTAACCTTTCCTACTCGGCTATGGCTACGTCAGCTACTCTGAAC 1370  
 QY 360 LysAspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsn 378  
 DB 1371 AAGCCGAAGTCAATCGAGGCTGTGGCGCTCAGGTCAACGCTACGACTCGTGAACCTT 1430  
 QY 379 AspValTyrTyrAlaPheGlnThrGlyAspPheValTrpPro-----AsnPheIle 396  
 DB 1431 GACATCAACCCCAACTCTCTCTCCACGCTGCTGGATGAAGCCCTACACCGCTCGTT 1490  
 QY 397 GluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleIleTyr---GlyAsp 415  
 DB 1491 CCGGACTCTGGAG-----CAGATCCCTGTC-----TGATCTATGCGCGTGTAT 1535  
 QY 416 AlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAsnIleTyrSer 435  
 DB 1536 GCTGATTTCATTTCAACTGGCTGGCAACAAGCCCTGGACTGAAGCCCTGGAAGTGGCC 1595  
 QY 436 GlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal----- 450  
 DB 1596 GGACAGCTGAATATGCTCCGCTGAGCTGGAGGATCTGGTCAITGTGCACATGAGCAC 1655  
 QY 451 AsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyr 470  
 DB 1656 ACGGCAAGAGATGCGCAGGTAAAGTCCCATGGCAACTTCACCTCAICGCTCTAT 1715  
 QY 471 GluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 490  
 DB 1716 GGTGTGGCCACATGTCGCGAIGGACCGCCGAGTCTCGAGTCTCGAGTTCTTCAACCGC 1775  
 QY 491 ThrIlePheGly---Trp 495  
 DB 1776 TGGTGGGAGGTGAATGG 1793

## RESULT 10

US-08-608-224-1  
 ; Sequence 1, Application US/08608224  
 ; Patent No. 5705376  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yaver, Debbie Sue  
 ; APPLICANT: Thompson, Sheryl Ann  
 ; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 5705376 No. 5705376disk of No. 5705376th America, Inc.  
 ; STREET: 405 Lexington Avenue, Suite 6400  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10174-6401  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

US-09-712-338-2\_COPY\_19\_555 (1-537) x US-08-608-224-1 (1-2068)

## RESULT 11

165 ProThrIleAspAlaThrLeuIleAaspThrThrAsnLeuAlaGluAlaIarPrpGlu 184  
Db : : : : :  
858 -----GTACAGCAGCACGGCTGCCTCGTCAAGAACGACGACTCTATGCC 896  
Qy 185 IleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAaspArgValGlnSerLysAsp 204  
Db : : : : :  
897 TTGCTTTACCCTCTCTTCACAAACAATTCCCG-----GAGTAGCTGAAGCAGGAC 944  
Qy 205 PheSerLeuTrpThrGluSerTyrglyGlyHisTyrglyProAlaPheAsnHisPhe 224  
Db : : : : :  
945 TTCACATTCCGGTGTAATCTATGCTGGTCACTATATCCCGCTCTCGCTTCGGAGATC 1004  
Qy 225 TyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsn 244  
Db : : : : :  
1005 CTGTCTCACAGAAGACGC-----AACATCAACCTGCGAC 1037  
Qy 245 SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTrpGlu 264  
Db : : : : :  
1038 TCGCTTCTCATTTGGCACGGCTTCACCGAGGATAACCCAGTAGTACGAGTACACCGTCCC 1099  
Qy 265 PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGlnThrValTyrAsnTyrMet 284  
Db : : : : :  
1098 ATGGCTCGCGGTGACGGCGGTTACCCAGCTGTCTTGACGACGAGCTCCCTGCCAGTCCATG 1157  
Qy 285 LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr 304  
Db : : : : :  
1158 GAC-----AACGCTCTCTCTCG-----TGCAGCTCATGATTGAGTCTTGCTACAGTTC 1208  
Qy 305 AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetycysArgAsp 324  
Db : : : : :  
1209 GAGAGCGCT-----TGGGTTGTGTCCGGSCCTCCACTACTGTAACAAC 1253  
Qy 325 AsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisPro 344  
Db : : : : :  
1254 GCCCTCTCTGCCCTTACCAACGC-----ACTGGCAGAACGCTCATGATGCCGTGTGAAG 1310  
Qy 345 TyrAspAspPro-----ThrProSerSerTyrTyrAsnLysPheLeuAla 359  
Db : : : : :  
1311 TGCAGAGTAGCTTAACCTTGCTACTCGGCTTGGCTACGCTACGCGACTACCTGAAC 1370  
Qy 360 LysAspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsn 378  
Db : : : : :  
1371 AAGCCCCAAGTCAFCGAGGCTGTGGCGCTGAGGCTCAACGGCTACGACTCGTCCAACCTT 1430  
Qy 379 AspValTyrTyrAlaPheGlnThrGlyAspPheValTrpPro-----AsnPheIle 396  
Db : : : : :  
1431 GACATCAACCGCAACTCTCTCTCCACGGTGACTGGNIGAAGCCCTACCAACCCCTCGTT 1490  
Qy 397 GluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAsp 415  
Db : : : : :  
1491 CCGGGACTCCCGGAG-----CAGATCCCTGTC-----TTGATCTATGCGGTGAT 1535  
Qy 416 AlaAspTyrIleCysAsnTrpPheGlyGlyGluAlaValSerLeuAlaAlaAsnTyrSer 435  
Db : : : : :  
1536 GCTGATTTTCATTTCGAATCGCTGGGCAACAAAGCCCTGAGCCCTGGAGTGGGCC 1595  
Qy 436 GlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----450  
Db : : : : :  
1596 GGACAGCTGNAATAGCTTCGGCTGAGCTGGAGGATCTGGTCACTGCGACAATGAGCAC 1655  
Qy 451 AsnGlyValGluTyrGlyuThrArgGluTyrGlyAsnPheSerPheThrArgValTyr 470  
Db : : : : :  
1656 ACGGGCAAGAAGATGGCCAGGTTTAAGTCCCATGGCAACTTCACCTTCATGCCGTCTCAT 1715  
Qy 471 GluAlaGlyHisGluValProTyrTyrGlyProIleAlaSerLeuGlnLeuPheAsnArg 490  
Db : : : : :  
1716 GTTGGTGGCCACATGGTCCCGATGGACCGCCGAGTCGAGTCTCGAGTCTTCTCAACCGC 1775  
Qy 491 ThrIlePheGly---Trp 495  
Db : : : : :  
1776 TGGTGGGAGCTGAATGG 1793  
Db

```

US-08-967-149-1
; Sequence 1, Application US/08967149
; Patent No. 593305
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5933050 No. 593305disk of No. 593305th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,149
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,452
; FILING DATE: 28-FEB-1996
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 572..632
; NAME/KEY: CDS
; LOCATION: join (571..633)
; US-08-967-149-1

Alignment Scores:
Pred. No.: 2,14e-55 Length: 2068
Score: 594.00 Matches: 155
Percent Similarity: 46.44% Conservative: 80
Best Local Similarity: 30.63% Mismatches: 174
Query Match: 20.57% Indels: 98
DB: 2 Gaps: 18

US-09-712-338-2_COPY_19_555 (1-537) x US-08-967-149-1 (1-2068)
QY 27 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46
Db 457 GCCTATGATCTCAGGCTCAAGAACCGATCTCTGGCTCTCTGGCATC-----GAC 507
QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGlu----- 61
Db 508 CCGCGCGTGAACAGTACACCGGTATCTCGATGACACAGA-GAATGATAAGCATTTGTT 566
QY 62 -----SerHisThrPhePhe----- 66

```

```

567 CTAGTAAGCACACACTTGGTTCAAGATCACCGCTTTTATATGCTCTGGATATCTAACGCA 626
QY 67 -----TrpPheGluAlaAArgHisAsnProGluThrAlaProIleThrLeuTrpLeu 84
Db 627 ACTTAGTGGTTCTTCGAGTCTCGCAATCACCCGAGAAATGATCCGGTTCTTCTGGCTG 686
QY 85 AsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluGlyProCysHis 104
Db 687 AACGGTGGCCCTGGGTCTCTCCCTCATCGGCTCTCTCATGAGAGCTTGGCCCTAGCAGC 746
QY 105 ValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu 124
Db 747 ATCAACAAGAAGATCCAGCCGCTCTACATGACTACGCTTGGAACTCCAACGCCCTCGTG 806
QY 125 LeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySer 144
Db 807 ATCTTCCTTGACCACGCTGTCAATGTCGGTTACTCTCTACATACTAATCTGCT----- 857
QY 145 IleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyr 164
Db 857 ----- 857
QY 165 ProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaTrpGlu 184
Db 858 -----GTACGACGACAGCTCGCTGCTGCCAAGGAGGCTCTATGCC 896
QY 185 IleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAsp 204
Db 897 TTGCTTACCCTCTCTTCAACAATCCCC-----GAGTATGCTAACGACGAC 944
QY 205 PheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsnHisPhe 224
Db 945 ITCCACATTCGCGGTGAATCTTATGCTGGTCACTATATCCCGCTCTTCGCTCGAGATC 1004
QY 225 TyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsn 244
Db 1005 CTGCTCTCACAAAGAGCC-----ACATCAACACTCTCAG 1037
QY 245 SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGlu 264
Db 1038 TCCGTTCTCATTTGGCAACGCTCTCACCGACGATACACCCAGTACGAGTACTACCGTCC 1097
QY 265 PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMet 284
Db 1098 ATGGCTTGGCGGTGACGCGGCTTACCCAGCTGTCTTGGACGAGAGCTCTGCCAGTCCATG 1157
QY 285 LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr 304
Db 1158 GAC-----AACGCTCTCTCTCGC---TGCAGCTCTATGATGAGTCTGTACAGTTCC 1208
QY 305 AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp 324
Db 1209 GAGAGCGCT-----TGGGTTGTGTCCCGCTCCCTACTCTACTGTAAACAAC 1253
QY 325 AsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisPro 344
Db 1254 GCCCTCTTGGCCCTTACCAGCGC---ACTGGCAGAACGCTCTATGATGCTGGTGAAG 1310
QY 345 TyrAspAspPro-----ThrProProSerTyrTyrAsnLysPheLeuAla 359
Db 1311 TGGCAGGATAGCTCTAACCTTCTCTCGGTATGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1370
QY 360 LysAspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsn 378
Db 1371 AAGCCCAAGTCTATCAGAGCTGTGGCGTGAAGTCAACGCTACCGCTACGCTGCTGCTGCT 1430
QY 379 AspValTyrTyrAlaPheGlnGlnThrGlyAspPheValTrpPro-----AsnPheIle 396
Db 1431 GACATCAACCGCAACTCTCTTCCAGGTGAGTGTGATGAAGCCCTACCAACGCCCTCGTT 1490
QY 397 GluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAsp 415

```

Db 1491 CCGGACTCCTGGAG-----CAGATCCCTGTC-----TTGATCATGCCGGTGAT 1535  
QY 416 AlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnIyrSer 435  
Db 1536 GCTGATTTTCATTGCAACATGGCTGGCGACACAGGCGCTGGACTGAAGCCCTGGAGTGGCC 1595  
QY 436 GlnAlaAlaGlnPheArgSerGlyAlaGlyTyrThrProLeuLysVal-----450  
Db 1596 GGACAGCTCGAATATCCCTCGCTGAGCTGGAGGATCTGTCATTGTCGACATGAGCAC 1655  
QY 451 AsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyr 470  
Db 1656 ACGGCAAGAAGATTGGCAGGTAAAGTCCCATGCCAACTTCACCTTCATGCGCTCTCTAT 1715  
QY 471 GluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 490  
Db 1716 GTGGTGGCCCATGTCCTCGATGACCGACCGAGCCGCGAGTCCGAGTCTCGATCTCAACCGC 1775  
QY 491 ThrIlePheGly---Trp 495  
Db 1776 TGGTGGGAGTGAATGG 1793

RESULT 12  
US-09-640-305-3  
Sequence 3, Application US/09640305  
Patent No. RE37447  
GENERAL INFORMATION:  
APPLICANT: Fleer, Reinhard  
Fournier, Alain  
Yen, Patrice  
TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR  
PREPARATION AND USE  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Rd. 3C43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19002  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/640,305  
FILING DATE: 16-Aug-2000  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,673  
FILING DATE: 06-FEB-1995  
APPLICATION NUMBER: WO PCT/FR93/00623  
FILING DATE: 23-JUN-1993  
APPLICATION NUMBER: FR 92/07785  
FILING DATE: 25-JUN-1992  
APPROPRIATE/AGENT INFORMATION:  
NAME: Smith, Julie K.  
REGISTRATION NUMBER: 38,619  
REFERENCE/DOCKET NUMBER: ST92040-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610)454-3839  
TELEFAX: (610)454-3808  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2503 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Kluveromyces lactis  
FEATURE:  
NAME/KEY: CDS

LOCATION: 387..1862  
OTHER INFORMATION: /product= "K. lactis protease C  
gene"  
/gene= "K1.PRC1"  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-640-305-3

Alignment Scores:  
Pred. No.: 1-87e-52 Length: 2503 -  
Score: 568.50 Matches: 151  
Percent Similarity: 44.65% Conservative: 66  
Best Local Similarity: 31.07% Mismatches: 184  
Query Match: 19.69% Indels: 85  
DB: 1 Gaps: 14

US-09-712-338-2\_COPY\_19\_555 (1-537) x US-09-640-305-3 (1-2503)

QY 27 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46  
Db 570 GCTTATTCATTGAGAAATAAACCTTTGGATCCCAATCTCTGGCGTT-----617  
QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePhe 66  
Db 618 GATACCGTGAACAATGGTCGGATATTAGATTACAGGACTCAAAACACTCTTTTAT 677  
QY 67 TrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGly 86  
Db 678 TGGTTTTTTCAGTCTAGAAATGACCCAGAGAAATGACCCAGTACTACTATGTTAAACGGT 737  
QY 87 GlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsn 106  
Db 738 GGTCTGGCTGTTCTCTCTTCGGTGGTCTTTTTCGATTTGGGACCTTCTCTATAGGA 797  
QY 107 SerThrPheAspTyrIleAsnProHisSerTyrAsnGluValSerAsnLeuPhe 126  
Db 798 GCTGATTTGAACCACTTATACCCCTACTCTTGGATTCACACGCTCTCTGTATATTC 857  
QY 127 LeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsn 146  
Db 858 CTAGATCAGCGCTGTGGTGGTGTCTCTCATACCGTGAC-----896  
QY 147 ProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThr 166  
Db 896 -----896  
QY 167 IleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeu 186  
Db 897 -----TCTAAGTGTCTACTACAGATGACCTGCCAAAGAGCTTTACATATTCTTA 947  
QY 187 GlnGlyPheLeuSerGlyLeuProSerIleAspSerArgValGlnSerLysAspPheSer 206  
Db 948 GATTTGTTCTTTGAAGATTCCCTCATTTG-----AGAAATAGCATTTCCAT 995  
QY 207 LeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGlu 226  
Db 996 ATCTCGGTGAATCATACGCGGTCTATTATTTCCAGATGCTCAT-----1043  
QY 227 GlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeu 246  
Db 1044 -----GAGATTGCTGTAGTGCATGCTGAGGATCTCTCAATCTATCTCATGTA 1094  
QY 247 GlyIleLeuAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTrpProGluPheAla 266  
Db 1095 TTAATTGGAAATGGATTACTGACCCACTGACTCAATACCAATATTACGACCGCATGCC 1154  
QY 267 ValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPhe 286  
Db 1155 TGTGGTGAAGGT---GGTATCCAGCGGTGTGGAA-----1187  
QY 287 AlaAsnGlnMetProAsnGlyCysGlnAsp-----296  
Db 1188 -----CCGGAAGATTGCTAGATATGAATAGTAGAATCTACCTCTATGCTATCG 1235

```

QY 297 LeuIleSerThrCysLysGluThrAsnArgThrAlaLeuAlaAspIleAlaLeuCysAla 316
DB 1236 CTTGTGGACCGCTGTACAAAGTCCCAT-----TCGTGTTTCTCTGTGTG 1280
QY 317 GluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArg 336
DB 1281 TTGGCTGACCGTATPTGTGAACACAGATTAATCTGGG---GTTTATGAGAAATCAGTAGG 1337
QY 337 GlyValTyrAspIleArg-----HisProTyrAspAspPro-----Thr 349
DB 1338 AACCTTACGATATTAGATCTAAGTCTCAAGCAGAGGATGATTCGGGTCCGCTGTATACAG 1397
QY 350 ProProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyVal 369
DB 1398 GAAGAAATTTATATCTCTGATTAATCTGAATCAGGAGGAAGTTCAGAGACGTTAGGACT 1457
QY 370 AsnIleAsnTyrThrGlnSer---AsnAsnAspValTyrTyrAlaPheGlnGlyThrGly 388
DB 1458 GATGTGAGTCTTCCCAAGGTGTAGCTCGGATCGGTATCGGTATCGGTATCGGTATCGGT 1517
QY 389 AspPheValTyrProAsnPheIleGluAspLeuGluGlyLeuAlaLeuAlaLeuProValArg 408
DB 1518 GATGACCGAGCCCA---TTCACACAGTACGTCGACAGATCTCTGATCAAGATATCAAT 1574
QY 409 ValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTyrPheGlyGlyGluAlaVal 428
DB 1575 GTCTTGATATATGACGAGCGATGAAGGATTAATTTGTAATTCGCTAGGAAATCTCGCTGG 1634
QY 429 SerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeu 448
DB 1635 ACTGAAANNTGGATGGAGGTATAACGAGAGAGTATAAANAACAAGTTTTCGAGACITGG 1694
QY 449 LysValAsnGlyValGlu-----TyrGlyGluThrArgGluTyrGlyAsnPheSerPhe 466
DB 1695 AAGAGTGAAGAAACAGATGACACATTTGGCGAACCACCAATCTTAAGCCCGCTAACCTAC 1754
QY 467 ThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGln 486
DB 1755 TTGAGAAATCATGATGCTGGACACATGTTCTTCACGACCAACCTGAAATTCATTACAA 1814
QY 487 LeuPheAsnArgThrIle 492
DB 1815 ATGGTGAATTCATGGATT 1832

RESULT 13
US-08-360-673-3
; Sequence 3, Application US/08360673
; Patent No. 5679544
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; APPLICANT: Fournier, Alain
; APPLICANT: Yeh, Patrice
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
; NUMBER OF INVENTION: PREPARATION AND USE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3043
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,673
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR93/00623

```

```

; FILING DATE: 23-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/07785
; FILING DATE: 25-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Julie K.
; REGISTRATION NUMBER: 38,619
; REFERENCE/DOCKET NUMBER: ST92040-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610)454-3839
; TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2503 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Kluyveromyces lactis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 387..1862
; OTHER INFORMATION: /product= "K. lactis protease C
; OTHER INFORMATION: /gene= "kl.PRC1"
; OTHER INFORMATION: /gene= "kl.PRC1"
US-08-360-673-3

```

```

Alignment Scores:
Pred. No.: 1-87e-52 Length: 2503
Score: 568.50 Matches: 151
Percent Similarity: 44.65% Conservative: 66
Best Local Similarity: 31.07% Mismatches: 184
Query Match: 19.69% Indels: 85
DB: 1 Gaps: 14

US-09-712-338-2_COPY_19_555 (1-537) x US-08-360-673-3 (1-2503)
QY 27 AlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrPhe 46
DB 570 GCTTATTCATTGAGAAATAAACCTTTGGATCCCAAACTCTCTGGCTT----- 617
QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePhe 66
DB 618 GATACCGTGAACAAAGTTCGGGATATTAGATACCAGGACTCAAAACACTCTTTTAT 677
QY 67 TrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTyrLeuAsnGly 86
DB 678 IGGTTTTTTCAGTCTAGAAATGACCCAGAGATGACCCAGGATGATGTTTAAACGGT 737
QY 87 GlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsn 106
DB 738 GGTCCTGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 797
QY 107 SerThrPheAspAspTyrIleAsnProHisSerTyrPheAsnGluValSerAsnLeuPhe 126
DB 798 CGTGAATTTGAACCCCAATTTATAACCCCTACTCTGGAATTCACACGCTCTGATATTC 857
QY 127 LeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsn 146
DB 858 CTAGATCAGCTGTGGTGTGGTGTCTCTCAACGGTGAC----- 896
QY 147 ProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThr 166
DB 896 ----- 896
QY 167 IleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaThrPheGluLeu 186
DB 897 -----TCTAAAGTGTCTACACAGATGAGCTGCCAAGACGTTTACATATCTTA 947
QY 187 GlnGlyPheLeuSerGlyLeuProSerIleAspSerArgValGlnSerLysAspPheSer 206
DB 948 GATTGTCTTTGAAAGATTCCTCATTTG-----AGAAATAACGATTTCAT 995

```

```

QY 207 LeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsnHisPheTyrGlu 226
Db   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
996 ATCTCGGTGAATCATACCGGTGCTATTTATTCACCAAGATTCCTCAAT----- 1043
QY 227 GlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnSheAsnSerLeu 246
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1044 -----GAGATTGCTGTAGTGCATGCTGAGGATTCCTCCATCAATCATCTATGTCAGTA 1094
QY 247 GlyIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAla 266
Db   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1095 TTAATTGGAATGATTCAGTCCACTGACTCAATACCAATATTACGAGCCGATGGCC 1154
QY 267 ValAsnAsnThrTyrGlyIleIleAlaValAsnGluThrValTyrAsnTyrMetLysPhe 286
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1155 TGTTGGTGAAGGT---GGTTATCAGCGGTGTTGAA----- 1187
QY 287 AlaAsnGlnMetProAsnGlyCysGlnAsp----- 296
Db   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1188 -----CCGGAAGATTCCTTAGATATGAATAGGAATCTACCTCTATGCTATCG 1235
QY 297 LeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAla 316
Db   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1236 CTGTGACGCGGTGTACAAAGTCCCAT-----TCTGTTTCTCTGTGTG 1280
QY 317 GluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArg 336
Db   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1281 TTGGCTACCGGTATTGTGACACAGATTACTGGG---GTTTATGAGAAATCAGTAGG 1337
QY 337 GlyValTyrAspIleArg-----HisProTyrAspAspPro-----Thr 349
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1338 AACCTTACCATATTAGATCTAAGTGTGAAGACAGAGATGATCCGGTGGCTGTTATCAG 1397
QY 350 ProProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyVal 369
Db   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1398 GAAGAAATTTATPCTCTGATTACTTGAATCAGAGGAGGATTCAGAGCCTTAGGACT 1457
QY 370 AsnIleAsnTyrThrGlnSer---AsnAsnAspValTyrTyrAlaPheGlnGluThrGly 388
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1458 GATGTGAGTCTTTCCAGGTTGTAGCTCGGATCGGTATCGGTTTCGCATTCACCTGGC 1517
QY 389 AspPheValTrpProAsnPheIleGluAspLeuGluIleLeuAlaLeuProValArg 408
Db   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1518 GATGACGAGGCCCA---TTCCACACAGTACGTCGACAACTCTTGATCAAGATAATCAAT 1574
QY 409 ValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaVal 428
Db   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1575 GTCCTGATATATGACGCGATAGGATTTATTTGTTATTTGCTAGGAAATCTCGCTGG 1634
QY 429 SerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGlyTyrProLeu 448
Db   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1635 ACTGAAATTTGGAATGGAGGTATACGAGAGATATAAACAAGTTTTCAGAACTGG 1694
QY 449 LysValAsnGlyValGlu-----TyrGlyGluThrArgGluTyrGlyAsnSerPhe 466
Db   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1695 AAGAGTGAAGAAACAGATAGACCATTTGGCGAAACCAAAATCTTATGGCCCTAACCTAC 1754
QY 467 ThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGln 486
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1755 TTGAGATCTATGATGCGGACATGATGTTCTCAGCACCACTGAAATTCATTACAA 1814
QY 487 LeuPheAsnArgThrIle 492
Db   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1815 ATGTGTAATTCATGGAAT 1832

RESULT 14
US-08-899-324-32
; Sequence 32, Application US/08899324
; Patent No. 5945329
; GENERAL INFORMATION:
; APPLICANT: Bredam, Klaus
; APPLICANT: Keilland-Brandt, Morten
; APPLICANT: Mortensen, Uffe

```

```

; APPLICANT: Olesen, Kjeld
; APPLICANT: Stennicke, Henning
; APPLICANT: Wagner, Fred
; TITLE OF INVENTION: CUSTOMIZED PROTEASES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS: Gould, Smith, Edell, Welter & Schmidt
; ADDRESSEE: Merchant, 3100 No. 5945329west Center, 90 S. 7th Street
; STREET: 3100 No. 5945329west Center, 90 S. 7th Street
; CITY: Minneapolis
; STATE: MN
; COUNTRY: U.S.A.
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,324
; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/329,892
; FILING DATE: 27-OCT-1994
; APPLICATION NUMBER: 08/144,704
; FILING DATE: 28-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettleberger, Denise M
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 8648.44USC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/332-5300
; TELEFAX: 612/332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2632 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 696..2291
; OTHER INFORMATION:
; US-08-899-324-32

Alignment Scores:
Pred. No.: 1,85e-47 Length: 2632
Score: 523.50 Matches: 147
Percent Similarity: 43.68% Conservative: 74
Best Local Similarity: 29.05% Mismatches: 200
Query Match: 18.13% Indels: 85
DB: 2 Gaps: 19

US-09-712-338-2_COPY_19_555 (1-537) x US-08-899-324-32 (1-2632)

QY 15 ProLysAsnProThrGlyValIlyThr-----LeuThrThrAlaAsnAsn 29
Db 939 CCAAAATTCCTCGAAGCAATCAAAACGAAAGAACACIGGGACTTTGTGTCAAGATGAC 998
QY 30 ValThrIleArgTyr-----LysGluProGlyAlaGluGlyVal 42
Db 999 GCAATGAAACATACAGCTTCGTCAACACAGATTAGACCCCTAAATCCTGGGCATT 1058
QY 43 CysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerProGlu--- 61
Db 1059 -----GACCCAAATGTCTACAGATACAGGGTTACTTGGATGTGGAGACGAGGAC 1109

```



QY 62 SerHisThrPhePheThrppheGluAlaArgHisAsnProGluThrAlaProIleThr 81  
 Db 1110 AAGCAITTCCTCTTTGGACATTTTGAAGTAGAACAGATCTGCAAGGATCCGGTCAATC 1169  
 QY 82 LeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGly 101  
 Db 1170 CTTTGGTTGAACGGGGTCCAGGTGTCTCTCACTAACCGGGCTGTCTTGAATTAGGA 1229  
 QY 102 ProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluVal 121  
 Db 1230 CCTCATCCATTTGGACCTGATTTGAACCCATCGGGAACCTTACTCTTGAACAGCAAT 1289  
 QY 122 SerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrVal 141  
 Db 1290 GCCACCGTGATCTCTCTTGACAGCCGTGCAACGTTCGGTTCGTATCC----- 1340  
 QY 142 AspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGln 161  
 Db 1341 -----GGTCTCTCA 1349  
 QY 162 GlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla 181  
 Db 1350 GGT-----GTTCACACTGCGCGCTGGTAAGAT 1382  
 QY 182 AlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGln 201  
 Db 1383 GTCTATAACTCTTGGAGTGTCTTCGATCAGTCCCTGAATACGTCACAGGCGCAA 1442  
 QY 202 SerLysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePhe 221  
 Db 1443 -----GATTTCACATCGCTGGGAATCTACGCGGCATACATCCCTGTTTGGC 1496  
 QY 222 AsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeu 241  
 Db 1497 TCGAATAATTGCTCTCACAGGACAC----- 1523  
 QY 242 AsnPheAsnSerLeuGlyIleIle-----AsnGlyIleIleAspGluAlaIleGlnAla 259  
 Db 1524 AACTTCAACTTACCTCCGCTCTTGTGATCGGAATGGCTCACTGACCCATGACTCAGTAT 1583  
 QY 260 ProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleIleValAsnGluThr 279  
 Db 1584 AACTATTACCAACCATGGCTGTGTGAAGGT---GGCGAACCATCTGT----- 1631  
 QY 280 ValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSer 299  
 Db 1632 -----ITGCCCTCGAGGATGCTCTCTATGAGACTCTTTGGAA 1673  
 QY 300 ThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu-----CysAlaGlu 317  
 Db 1674 CGTTGTTGGGCTGTATCGAGTCGTGTATGACTCGCAATGGTCTGTCTGTGTCCA 1733  
 QY 318 AlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGly 337  
 Db 1734 GTACCATTTATGTAAATACGCCCAATGGCTCTTACCAAGT---ACCGGCAAAAC 1790  
 QY 338 ValTyrAspIleArgHisProTyrAspAsp-----ProThrProProSer 352  
 Db 1791 GTTTAGGATATCAGGAAGGATTTGGAAGTGGCAATTTGTGCTACCCACAGTTACAGAT 1850  
 QY 353 TyrTyrAsnLysPheLeuAlaLysPheSerValMetAspAlaIleGlyValAsnIleAsn 372  
 Db 1851 ----ATCGACGACTACTTAACACGAGTACTAGCAAGAGTGTCTCGGGTGTATGGATG 1907  
 QY 373 TyrThrGluSer---AsnAsnAspValTyrTyrAlaPheGlnGlnThrGlyAspPheVal 391  
 Db 1908 CACTAGCAATCTGTAACTTCGATATCAACAGAAATTTCTGTTCGCGGTGATGGATG 1967  
 QY 392 TrpProAsnPheIleGluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeu 411  
 Db 1968 AAGCCT---TACCACACCGCCGTAAACAGATCTTTGAATCAACAGCTACCCATCTGGTA 2024  
 QY 412 IleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAla 431

Db 2025 TATCAGCGCGAARAGATTTCTCTGTAACTGGTGGTAATAAGCGGTGAGCGATGTC 2084  
 QY 432 AlaAsnTyrSerGlnAlaAlaGlnPheArgSer-----AlaGlyTyrThrProLeu 448  
 Db 2085 TTGCCATGGAAGTACGAGGAAGATTTCAGACCAAAAGTACGTAACCTGACCTGCTTCT 2144  
 QY 449 LysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArg 468  
 Db 2145 ATCACCAGCAGAGTCCGCT---GGTAGCTCAATCTCAAGCAGCTTCCCTATTGAGA 2201  
 QY 469 ValTyrGluAlaGlyHisGluValPheProTyrTyrGlnProIleAlaSerLeuGlnLeuPhe 488  
 Db 2202 GTCTTCAATGGTGGCCACATGTTCCATTGACGTCCCTGAAACGCCCTTAAGTATGTT 2261  
 QY 489 AsnArgThrIlePheGly 494  
 Db 2262 AACGAATGGATCCACGGT 2279  
 RESULT 15  
 US-08-329-892B-32  
 ; Sequence 32, Application US/08329892B  
 ; Patent No. 6187579  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bredam, Klaus  
 ; APPLICANT: Keiland-Brandt, Morten  
 ; APPLICANT: Mortensen, Uffe  
 ; APPLICANT: Olesen, Kjeld  
 ; APPLICANT: Stennicke, Henning  
 ; APPLICANT: Wagner, Fred  
 ; TITLE OF INVENTION: CUSTOMIZED PROTEASE  
 ; NUMBER OF SEQUENCES: 33  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
 ; STREET: 3100 No. 6187579west Center, 90 S. 7th Street  
 ; CITY: Minneapolis  
 ; STATE: MN  
 ; COUNTRY: U.S.A.  
 ; ZIP: 55402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq Version 1.5  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/329,892B  
 ; FILING DATE: 27-OCT-1994  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/144,704  
 ; FILING DATE: 28-OCT-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kettleberger, Denise M  
 ; REGISTRATION NUMBER:  
 ; REFERENCE/DOCKET NUMBER: 8648.44US01  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 612/332-5300  
 ; TELEFAX: 612/332-9081  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 32:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2632 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; FRAGMENT TYPE:  
 ; ORIGINAL SOURCE:  
 ; FEATURE:  
 ; NAME/KEY: Coding Sequence  
 ; LOCATION: 696...2291

# OTHER INFORMATION:

US-08-329-892B-32

## Alignment Scores:

Pred. No.: 1,85e-47 Length: 2632  
 Score: 523.50 Matches: 147  
 Percent Similarity: 43.68% Conservative: 74  
 Best Local Similarity: 29.05% Mismatches: 200  
 Query Match: 18.13% Indels: 85  
 Gaps: 19

US-09-712-338-2\_COPY\_19\_555 (1-537) x US-08-329-892B-32 (1-2632)

QY 15 ProLysAsnProThrGlyValIleThr-----LeuThrThrAlaAsnAsn 29  
 DQ 939 CCAAAATCCCTCAACCAATCAACGAGAGACTGGGACTTTCTGGTCAAGANTGAC 998  
 QY 30 ValThrIleArgTyr-----LysGluProGlyAlaGluGlyVal 42  
 DQ 999 GCAATTTGAAAGACTATCAGCTTCGTGTCACACAGATTAAAGGACCCCTAAATTCCTGGGCATT 1058  
 QY 43 CysGluThrThrProGlyValIleThrSerGlyTyrValIleThrSerProGlu--- 61  
 DQ 1059 -----GACCCAAATGTACACAGTACACGGGTACTTGGATGTGGAAGCAGGAC 1109  
 QY 62 SerHisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIleThr 81  
 DQ 1110 AAGCATTTCTCTTTGGACTTTTGAAGTAGAAGCATCTCTCAAGAGATCGGTATC 1169  
 QY 82 LeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGly 101  
 DQ 1170 CTTTGGTTGAACGGGGTCCAGGTGTTCTTCTACTAACCGGGCTGTTCTTTGAATTAGGA 1229  
 QY 102 ProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluVal 121  
 DQ 1230 CCCTATFCAATCGACTGATTTGAACCCATCGGAAACCCCTTACTCTTGGAAACAGCAAT 1289  
 QY 122 SerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTrpAspThrVal 141  
 DQ 1290 GCACCGGTGATCTCTCTTGACAGAGCTGACAGCTGTCAGCTTCTGATTC 1340  
 QY 142 AspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGln 161  
 DQ 1341 -----GGGTCTCTCA 1349  
 QY 162 GlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAla 181  
 DQ 1350 GGI-----GTTTCCACACTGTCGCCGCTGGTAAGAT 1382  
 QY 182 AlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGln 201  
 DQ 1383 GTCTATAAATCTCTGGAGTTGTTCTTCGATCAGTTCCCTGAATACGTCACAAAGGCGCAA 1442  
 QY 202 SerIlyAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePhe 221  
 DQ 1443 -----GATTTCCACATCGCTGGGGAATCTACGCGCGCATTAACATCCCTGTTTTCGC 1496  
 QY 222 AsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeu 241  
 DQ 1497 TCTGAATTTGTCTCACAGGACAGA----- 1523  
 QY 242 AsnPheAsnSerLeuGlyIleIle-----AsnGlyIleIleAspGluAlaIleGlnAla 259  
 DQ 1524 AACTTCAACTTAACCTCGCTTCTGATCGGAATGGCCCTACTGACCCATTGACTCAGTAT 1583  
 QY 260 ProTyrTrpProGluPheAlaValAsnThrTyrGlyIleLysAlaValAsnGluThr 279  
 DQ 1584 AACTATTACGAACCAATGGCTGCGTGAAGGT---GGCGAACCATCTGTT----- 1631  
 QY 280 ValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSer 299  
 DQ 1632 -----TTGCCCTCGAGGAGTCTGCTATGGAAGACTCTTTTGAA 1673

QY 300 ThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu-----CysAlaGlu 317  
 DQ 1674 CGTTGTTGGGCTTGATCGAGTCTGCTATGACTCGCAATCGCTGCTGGTCTGTGTTCCA 1733  
 QY 318 AlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGly 337  
 DQ 1734 GTTACCATTTATTGTAATACGCCCAATTCGGTCTCTTACCAACGT---ACCGGACGANAAC 1790  
 QY 338 ValTyrAspIleArgHisProTyrAspAsp-----ProThrProProSer 352  
 DQ 1791 GTTACGATATCAGGAAGGATTGTGAAGGTGGCAATTTGTGCTACCCACAGTTTACAAGAT 1850  
 QY 353 TyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsn 372  
 DQ 1851 ---ATCGAGACTACTTAAACACGAGTACTCTCAAGAAGCTGTGCGTGGAGGTGAC 1907  
 QY 373 TyrThrGlnSer--AsnAsnAspValTyrTyrAlaPheGlnGlnThrGlyAspPheVal 391  
 DQ 1908 CACTACGAATCCCTGTACTTCGATATCAACAGAAATTTCTGTTTGGGGTGTATGGATG 1967  
 QY 392 TrpProAsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeu 411  
 DQ 1968 AAGCCT--TACCACACCGCGGTAAACAGATCTTTTGAATCAAGACCTACCCATTCTGTA 2024  
 QY 412 IleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAla 431  
 DQ 2025 TATGCAGCGCATAAAGATTTCTCTGTAACTGTGGTGTATTAAGCGCTGGACGGATGTC 2084  
 QY 432 AlaAsnTyrSerGlnAlaAlaGlnPheArgSer-----AlaGlyTyrThrProLeu 448  
 DQ 2085 TTGCCATGGGAAGTACGACGAGAAATTTGCAAGCCAAAGTACGTACTGACTGCTTCT 2144  
 QY 449 LysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArg 468  
 DQ 2145 ATCACCGACGAGGTCTGCT---GGTGAAGTCAATCTTCAAGACCTTACCTATTGAGA 2201  
 QY 469 ValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGluLeuPhe 488  
 DQ 2202 GTCTTCAATGGTGGCCACATGGTTCCTATTGACGTCCCTGAAACGCTTAAAGTATGTT 2261  
 QY 489 AsnArgThrIlePheGly 494  
 DQ 2262 ACGAATGGAATCCACGGT 2279

Search completed: September 17, 2003, 01:46:56

Job time : 152 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 22:47:57 : Search time 36 Seconds  
(without alignments)  
631.137 Million cell updates/sec

Title: US-09-712-338-2\_COPY\_19\_555  
Perfect score: 2887  
Sequence: 1 LPSGTASVGRRLPKNPTG.....HTQSSVPLPRAISMSSVGM 537

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCITUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match	Length				
1	2883	99.9	554	3	US-08-943-714-2		Sequence 2, Appli
2	894	31.0	423	3	US-08-943-714-9		Sequence 9, Appli
3	630	21.8	557	1	US-08-309-341-2		Sequence 2, Appli
4	630	21.8	557	1	US-08-608-267-2		Sequence 2, Appli
5	630	21.8	557	1	US-08-608-452-2		Sequence 2, Appli
6	630	21.8	557	1	US-08-608-224-2		Sequence 2, Appli
7	630	21.8	557	2	US-08-967-149-2		Sequence 2, Appli
8	623	21.6	557	1	US-08-309-341-4		Sequence 4, Appli
9	623	21.6	557	1	US-08-608-267-4		Sequence 4, Appli
10	623	21.6	557	1	US-08-608-452-4		Sequence 4, Appli
11	623	21.6	557	1	US-08-608-224-4		Sequence 4, Appli
12	623	21.6	557	2	US-08-967-149-4		Sequence 4, Appli
13	568.5	19.7	491	1	US-09-640-305-4		Sequence 4, Appli
14	568.5	19.7	491	1	US-08-360-673-4		Sequence 4, Appli
15	523.5	18.1	532	2	US-08-899-324-33		Sequence 33, Appli
16	523.5	18.1	532	3	US-08-329-892B-33		Sequence 33, Appli
17	513.5	17.8	421	2	US-08-807-263-4		Sequence 3, Appli
18	361	12.5	476	2	US-08-828-488-3		Sequence 3, Appli
19	361	12.5	476	4	US-09-299-689A-3		Sequence 3, Appli
20	348.5	12.1	477	2	US-08-828-488-1		Sequence 1, Appli
21	348.5	12.1	477	4	US-09-299-689A-1		Sequence 1, Appli
22	331.5	11.5	480	2	US-08-828-488-8		Sequence 1, Appli
23	331.5	11.5	480	4	US-09-299-689A-8		Sequence 8, Appli
24	331.5	11.5	480	4	US-09-702-705-336		Sequence 8, Appli
25	331.5	11.5	480	4	US-09-736-457-336		Sequence 336, App
26	312.5	10.8	471	2	US-08-828-488-7		Sequence 336, App
27	312.5	10.8	471	4	US-09-299-689A-7		Sequence 7, Appli

28	266.5	9.2	523	3	US-08-943-714-11	Sequence 11, Appli
29	246.5	8.5	446	1	US-08-665-966-10	Sequence 10, Appli
30	246.5	8.5	446	3	US-09-041-780-10	Sequence 10, Appli
31	208	7.2	530	3	US-08-943-714-12	Sequence 12, Appli
32	204.5	7.1	481	3	US-08-943-714-10	Sequence 10, Appli
33	193	6.7	35	3	US-08-943-714-5	Sequence 5, Appli
34	172	6.0	179	1	US-08-665-966-8	Sequence 8, Appli
35	172	6.0	179	3	US-09-041-780-8	Sequence 8, Appli
36	169.5	5.9	351	2	US-08-828-488-5	Sequence 5, Appli
37	169.5	5.9	351	4	US-09-299-689A-5	Sequence 5, Appli
38	142	4.9	29	3	US-08-943-714-6	Sequence 6, Appli
39	127.5	4.4	771	4	US-09-634-238-229	Sequence 229, App
40	127.5	4.4	771	4	US-09-634-238-262	Sequence 262, App
41	126	4.4	385	2	US-08-387-942C-24	Sequence 24, Appli
42	126	4.4	1403	2	US-08-387-942C-3	Sequence 3, Appli
43	121	4.2	618	3	US-09-199-290-34	Sequence 34, Appli
44	120	4.2	385	2	US-08-387-942C-27	Sequence 27, Appli
45	120	4.2	872	2	US-08-387-942C-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1  
US-08-943-714-2  
; Sequence 2, Application US/08943714  
; Patent No. 6187578  
; GENERAL INFORMATION: 2/13/01  
; APPLICANT: Blinksy, Alexander  
; APPLICANT: Berk, Randy  
; APPLICANT: Rey, Michael  
; APPLICANT: Gollightly, Elizabeth  
; APPLICANT: Klotz, Alan  
; APPLICANT: Mathisen, Thomas Erik  
; APPLICANT: Dambmann, Claus  
; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS: Encoding Same  
; ADDRESS: No. 61875780 No. 6187578disk of No. 6187578th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,714  
FILING DATE: 03-OCT-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENCY INFORMATION:  
NAME: Lambiris, Elias J  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4990.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 554 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-943-714-2

Query Match 99.9%; Score 2883; DB 3; Length 554;

Best Local Similarity 100.0%; Pred. No. 1.2e-255;  
Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFGSTPASVGRQLPKNPQVTKLTATANNVTTRYKEPGAEGVCEITPGVKSYSGYDTP 60  
Db 19 LFGSTPASVGRQLPKNPQVTKLTATANNVTTRYKEPGAEGVCEITPGVKSYSGYDTP 78

QY 61 ESHTFFWFEARHNPETATITLWNGPGSDSLIGLFEELGPHVNSTFDDYINPWSNE 120  
Db 79 ESHTFFWFEARHNPETATITLWNGPGSDSLIGLFEELGPHVNSTFDDYINPWSNE 138

QY 121 VSNLLFSLQPLGVGSYSYDTPGSIINPVTGVVENSFFAGVQGRYPTIDATLIDTTLAAE 180  
Db 139 VSNLLFSLQPLGVGSYSYDTPGSIINPVTGVVENSFFAGVQGRYPTIDATLIDTTLAAE 198

QY 181 AWEILQGLSLPLSDRSVQKDFSLWTESYGHYGPAPFNHFFQNERIANGSVNGVQ 240  
Db 199 AWEILQGLSLPLSDRSVQKDFSLWTESYGHYGPAPFNHFFQNERIANGSVNGVQ 258

QY 241 LNFNSLGIINGIIDEALQAPYEPFAVNNYIGKAYNETVYNTKMEANQMPNCQDLIST 300  
Db 259 LNFNSLGIINGIIDEALQAPYEPFAVNNYIGKAYNETVYNTKMEANQMPNCQDLIST 318

QY 301 CKQTNRTALADYALCAEATNMCRDNVEGYPYAFAGRGVYDIRHPYDDPTPPSYNKFAL 360  
Db 319 CKQTNRTALADYALCAEATNMCRDNVEGYPYAFAGRGVYDIRHPYDDPTPPSYNKFAL 378

QY 361 DSVMDALGVNINTQSNNDVYAFQOGDFVWPNFTEDEILALFVRVSLIYGDADYIC 420  
Db 379 DSVMDALGVNINTQSNNDVYAFQOGDFVWPNFTEDEILALFVRVSLIYGDADYIC 438

QY 421 NFWGGOAVSLAANSQAQFRSAGYTPLVKNGVEYGETREYGNFSTRVYEAGHEVPYQ 480  
Db 439 NFWGGOAVSLAANSQAQFRSAGYTPLVKNGVEYGETREYGNFSTRVYEAGHEVPYQ 498

QY 481 PIASQLQFNRTIFGWIDIAEQKKIWPSTKNGTATATHTQSSVPLPTATSMSSVGM 536  
Db 499 PIASQLQFNRTIFGWIDIAEQKKIWPSTKNGTATATHTQSSVPLPTATSMSSVGM 554

RESULT 2  
US-08-943-714-9  
; Sequence 9, Application US/08943714  
; Patent No. 6187578  
; GENERAL INFORMATION:  
; APPLICANT: Blinkovsky, Alexander  
; APPLICANT: Berk, Randy  
; APPLICANT: Rey, Michael  
; APPLICANT: Golightly, Elizabeth  
; APPLICANT: Klotz, Alan  
; APPLICANT: Mathisen, Thomas Erik  
; APPLICANT: Dammann, Claus  
; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids  
; TITLE OF INVENTION: Encoding Same  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 61875780 No. 61875780disk of No. 6187578th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/943,714  
; FILING DATE: 03-OCT-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J

REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4990.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 423 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-943-714-9

Query Match 31.0%; Score 894; DB 3; Length 423;  
Best Local Similarity 40.1%; Pred. No. 1.7e-73;  
Matches 186; Conservative 66; Mismatches 160; Indels 52; Gaps 8;

QY 41 GVCETTPGVKYSYGYDTPSPESHFFWFEARHNPETATITLWNGPGSDSLIGLFEEL 100  
Db 6 GICETTPGVKYSYGYDTPSPESHFFWFEARHNPETATITLWNGPGSDSLIGLFEEL 65

QY 101 GPGH-VNSTFDDYINPWSNEVSNLLFSLQPLGVGSYSYDTPGSIINPVTGVVENSFFAG 159  
Db 66 GPGH-VNSTFDDYINPWSNEVSNLLFSLQPLGVGSYSYDTPGSIINPVTGVVENSFFAG 104

QY 160 VQGRYPTIDATLIDTTLAAEAEILQGLSLPLSDRSVQKDFSLWTESYGHYGP 219  
Db 105 -----TDDVT---STVTAAPYVWNLQAFYAPQRYEESR---DFAITFESYGHYGP 151

QY 220 FNFHYQNERIANGSVNGVNLNLSGLIINGIIDEALQAPYEPFAVNNYIGKAYNET 279  
Db 152 FASYIEQQAAIKAGSVTQGNVNI VALGVNNGWIDSTIQERAYIDFSYNNSYQDIIDST 211

QY 280 VYNTKMEANQMPNCQDLISTCKQTNRTALADYALCAEATNMCRDNVEGYPYAFAGRGVY 339  
Db 212 RDSLLDAYN---NOCFLAQCCSSGST-----SDCINADSVCYQNTIEGPSSSDPDVY 263

QY 340 DIRHPYDDPTPPSYNKFALAKSDYMDALGVNINTQSNNDVYAFQOGDFVWPNFTEDE 399  
Db 264 DIREPSNDPYPKTYSTYLSPTVVKAIGARTNYQECFNGPYNKFASITGDNP-RSELTSL 322

QY 400 EEILALFVRVSLIYGDADYICNFWGGOAVSLAANSQAQFRSAGYTPLVKNGVEYGETR 459  
Db 323 SSVVQSGINVLWAGDADWICNLGNVEVANA VDFPQNAQFSALDAPLTYNGVEKGQFK 382

QY 460 EYGNFSTRVYEAGHEVPYQPIASQLQFNRTIFGWIDIAEQKK 503  
Db 383 TVDNFSLKLVYAGHEVPYQPIALQAFKQII-----QKK 418

RESULT 3  
US-08-309-341-2  
; Sequence 2, Application US/08309341  
; Patent No. 5594119  
; GENERAL INFORMATION:  
; APPLICANT: Yaver, Debbie Sue  
; APPLICANT: Thompson, Sheryl Ann  
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 55941190 No. 5594119disk of No. 5594119th America, Inc.  
; STREET: 405 Lexington Avenue, Suite 6400  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/309,341
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
;
US-08-309-341-2

Query Match          21.8%; Score 630; DB 1; Length 557;
Best Local Similarity 31.8%; Pred. No. 4.3e-49;
Matches 154; Conservative 81; Mismatches 174; Indels 76; Gaps 17;

QY 27 ANNVTIRYKPGAGVCGTTPGVKSYGYVDTP-ESHTFFWFEEARHNPTAPITLWLN 85
Db 132 AYDLRVKKTDPGSLGI---DPGVKQYTGILDDNENDKHLFYWFESRNDPDPVLWLN 188
QY 86 GPGSDSLIGLFEELGPGCHVNSTFDDYINPHSWNEVSNLLFLSOPLGVGFSYSTVDGSI 145
Db 189 GPGCCSLTGLFMELGPGSSINKTQPVNDYAWNSASVIFLDQPVNVGYYSNSA---- 244
QY 146 NPVTGVVENSFAGVQGRYPTIDATLIDTTNLAEEAAWEILQGLSLPSLDSRVQSKDF 205
Db 245 -----VSDIVAAGKDVYALLTLFFKQFP-----EYAKODF 274
QY 206 SLWTSTGGHYGPAFFNHFYEQNERIANGSVNGVQLNFSGLIINGIIDEAIQAPYPEF 265
Db 275 HIAGESYAGHYIPVFAFSEILSHKKR-----NINLQSVLIGNGLTDGTYQYVYRPM 325
QY 266 AVNNTYGIKAVNETVYVMKFNANPNCQDLSICQTKNTALADYALCAEATMCRDN 325
Db 326 ACDGGYPVAVLDESSCOSMD--NALPR-COSMIESCYSSESA-----WVCVPASIYCNNA 377
QY 326 VEGPYAFAGRGVYDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGVNLN-VTQSNND 379
Db 378 LLAPYQR-TGONVYDVRGKCEDSSNLCYSAMGYVSDYLNKPEVTEAVGAEVNGVDSNCFD 436
QY 380 VYFAFOQTGFVWP--NFTEDLEIILALPVRVSLIY-GDADYICNWFPGQAVSLAANSYQ 436
Db 437 INRNFLPHGDWMPYHRLVPGILLE--QIPV--LIVAGDADFICNWLGNKAWTEALEWPG 491
QY 437 AAQFRSAGYTPLAV-----NGVEYGETREYGNESFTFVYAGHEVPPYQPIASLQFNRT 491
Db 492 QAYASAELEDIVVDNEHTGKIGQVKSNGHETFMRLYGGGHVMPMDQPESSLEFNRW 551
QY 492 IFG-W 495
Db 552 LGGEW 556

RESULT 4
US-08-608-267-2
; Sequence 2, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Iaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5688663o No. 5688663disk of No. 5688663th America, Inc.

```

```

; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,267
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
;
US-08-608-267-2

```

```

Query Match          21.8%; Score 630; DB 1; Length 557;
Best Local Similarity 31.8%; Pred. No. 4.3e-49;
Matches 154; Conservative 81; Mismatches 174; Indels 76; Gaps 17;

QY 27 ANNVTIRYKPGAGVCGTTPGVKSYGYVDTP-ESHTFFWFEEARHNPTAPITLWLN 85
Db 132 AYDLRVKKTDPGSLGI---DPGVKQYTGILDDNENDKHLFYWFESRNDPDPVLWLN 188
QY 86 GPGSDSLIGLFEELGPGCHVNSTFDDYINPHSWNEVSNLLFLSOPLGVGFSYSTVDGSI 145
Db 189 GPGCCSLTGLFMELGPGSSINKTQPVNDYAWNSASVIFLDQPVNVGYYSNSA---- 244
QY 146 NPVTGVVENSFAGVQGRYPTIDATLIDTTNLAEEAAWEILQGLSLPSLDSRVQSKDF 205
Db 245 -----VSDIVAAGKDVYALLTLFFKQFP-----EYAKODF 274
QY 206 SLWTSTGGHYGPAFFNHFYEQNERIANGSVNGVQLNFSGLIINGIIDEAIQAPYPEF 265
Db 275 HIAGESYAGHYIPVFAFSEILSHKKR-----NINLQSVLIGNGLTDGTYQYVYRPM 325
QY 266 AVNNTYGIKAVNETVYVMKFNANPNCQDLSICQTKNTALADYALCAEATMCRDN 325
Db 326 ACDGGYPVAVLDESSCOSMD--NALPR-COSMIESCYSSESA-----WVCVPASIYCNNA 377
QY 326 VEGPYAFAGRGVYDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGVNLN-VTQSNND 379
Db 378 LLAPYQR-TGONVYDVRGKCEDSSNLCYSAMGYVSDYLNKPEVTEAVGAEVNGVDSNCFD 436
QY 380 VYFAFOQTGFVWP--NFTEDLEIILALPVRVSLIY-GDADYICNWFPGQAVSLAANSYQ 436
Db 437 INRNFLPHGDWMPYHRLVPGILLE--QIPV--LIVAGDADFICNWLGNKAWTEALEWPG 491
QY 437 AAQFRSAGYTPLAV-----NGVEYGETREYGNESFTFVYAGHEVPPYQPIASLQFNRT 491
Db 492 QAYASAELEDIVVDNEHTGKIGQVKSNGHETFMRLYGGGHVMPMDQPESSLEFNRW 551
QY 492 IFG-W 495
Db 552 LGGEW 556

```

Db 552 LGGEW 556

RESULT 5

US-08-608-452-2

; Sequence 2, Application us/08608452

; Patent No. 5693510

; GENERAL INFORMATION:

; APPLICANT: Yaver, Debbie Sue

; APPLICANT: Thompson, Sheryl Ann

; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5693510disk of No. 5693510disk of No. 5693510th America, Inc.

; STREET: 405 Lexington Avenue, Suite 6400

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/608,452

; FILING DATE: 28-FEB-1996

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/309,341

; FILING DATE: 20-SEP-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lowney, Karen A.

; REGISTRATION NUMBER: 31,274

; REFERENCE/DOCKET NUMBER: 4247.000-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 867 0123

; TELEFAX: 212 867 0298

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 557 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Aspergillus Niger

; US-08-608-452-2

Query Match 21.8%; Score 630; DB 1; Length 557;

Best Local Similarity 31.8%; Pred. No. 4.3e-49;

Matches 154; Conservative 81; Mismatches 174; Indels 76; Gaps 17;

QY 27 ANNVIRYKEGAGCVETTPGVKSYSGYVDTSP-ESHITFFWFPEARHNPEIAPITWLN 85

Db 132 AYDLRKKTDPSGLI---DPGVKQYTGVLDDNENDKHLFFWFESRNDPNDPVVLWN 188

QY 86 GPGGSDSLIGLFEELGCPCHVNSTDDYINPHSWNEVSNLLFLSQPLGVGFSYSDTVGSI 145

Db 189 GPGGCSLITGLFMELGPSINKKIQPVYNDYAWNNSAVIFLDQPVNVGYYSNSA--- 244

QY 146 NPVTGVVNSSFAGVQGRYPTIDATLIDTTNLAAPAAWELLOGFLSLPSLDSRVQSKDF 205

Db 245 -----VSDTVAAGKDVYALLTLFPKQPP-----EYAKQDF 274

QY 206 SLWTSYGGHYGPAFFENHYEONEELANGSVNGLNLSICLIINGIIDEALQAPYEF 265

Db 275 HTAGESYAGHYTPVFASEILSHKK-----NINQSVLIGNLGTLDTGYTQYTYRPM 325

QY 266 AVNNYTGIKAVNETVYNYMKFANOMPNGCQDLISCTKQTNRTALADYALCAEATNMCRDN 325

Db 326 ACGDGGYPVLDESSCOSMD--NALPR-CQSMIESCYSSESA-----WVCVPASTYCNA 377

QY 326 VEGPYAFAGRGVYDISHPYDDP-----TPPSYNNKEIAKDSVMDAIGVNLN-YTQSNND 379

Db 378 LLAPYQR-TGQNVYDVRGKCEDSSNLCSAMGYVSDYLNKPEVTEAYGAENVGYDSNED 436

QY 380 VYAFQQTGDFVWP--NFIEDLEILLALPVRSLLIY-GDADYIGNWFGGAQVSLAANYSQ 436

Db 437 INRNLEFHGDWMPKYHRLVPLLE--QIPV---LIYAGDADFICNWLGNKAWTALWPG 491

QY 437 AAOFSAGYTPLVK-----NGVEXGETREYGNFSTRVYEGAGHEVPYQPTIASLQLFNRT 491

Db 492 QAEVASAELELDLVIDNEHTKKGIGQVKSNGHGFEMRLYGGGHVMPDQPESSLEFENRW 551

QY 492 IFG-W 495

Db 552 LGGEW 556

RESULT 6

US-08-608-224-2

; Sequence 2, Application US/08608224

; Patent No. 5705376

; GENERAL INFORMATION:

; APPLICANT: Yaver, Debbie Sue

; APPLICANT: Thompson, Sheryl Ann

; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5705376 No. 5705376disk of No. 5705376th America, Inc.

; STREET: 405 Lexington Avenue, Suite 6400

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/608,224

; FILING DATE: 28-FEB-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/309,341

; FILING DATE: 20-SEP-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lowney, Karen A.

; REGISTRATION NUMBER: 31,274

; REFERENCE/DOCKET NUMBER: 4247.000-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 867 0123

; TELEFAX: 212 867 0298

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 557 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Aspergillus Niger

; US-08-608-224-2

Query Match 21.8%; Score 630; DB 1; Length 557;

Best Local Similarity 31.8%; Pred. No. 4.3e-49;

Matches 154; Conservative 81; Mismatches 174; Indels 76; Gaps 17;

QY 27 ANNVIRYKEGAGCVETTPGVKSYSGYVDTSP-ESHITFFWFPEARHNPEIAPITWLN 85

Db 132 AYDLRKKTDPSGLI---DPGVKQYTGVLDDNENDKHLFFWFESRNDPNDPVVLWN 188

QY 86 GPGGSDSLIGLFEELGCPCHVNSTDDYINPHSWNEVSNLLFLSQPLGVGFSYSDTVGSI 145

Db 189 GPGGCSLITGLFMELGPSINKKIQPVYNDYAWNNSAVIFLDQPVNVGYYSNSA--- 244

QY 146 NPVTGVVNSSFAGVQGRYPTIDATLIDTTNLAAPAAWELLOGFLSLPSLDSRVQSKDF 205

Db 245 -----VSDTVAAGKDVYALLTLFPKQPP-----EYAKQDF 274

QY 206 SLWTSYGGHYGPAFFENHYEONEELANGSVNGLNLSICLIINGIIDEALQAPYEF 265

Db 275 HTAGESYAGHYTPVFASEILSHKK-----NINQSVLIGNLGTLDTGYTQYTYRPM 325

QY 266 AVNNYTGIKAVNETVYNYMKFANOMPNGCQDLISCTKQTNRTALADYALCAEATNMCRDN 325

Db 326 ACGDGGYPVLDESSCOSMD--NALPR-CQSMIESCYSSESA-----WVCVPASTYCNA 377

```

Db 189 GPGCCSLTGLFMELGPFSSINKKIQPVYNDYAWNSNASVIFLDQPVNVGYYSNSA----- 244
QY 146 NPVTGVVENSFAGVQGRYPIDATLDTTNIAAEAAWEILQGLSLPSLDSRVQSKDF 205
Db 245 -----VSDIVAAGKDVYALLTLFFKOPF-----EYAKQDF 274
QY 206 SLWTESYGHYGFAPFHFYEQNRIANGSVNGVQLNFNSLGIINGIIDEAIAQIAPYPEF 265
Db 275 HIAGESYAGHYPIPVFASEILSHKKR-----NINLQSVLIGNGLIDGVTOYEYRPM 325
QY 266 AVNNTYGIKAVNETVYNYMKFANPONGCDLISTCKQTNRTALADYALCAEATNMCRDN 325
Db 326 ACGDGGYPAVLDESSCQSD--NALPR-COSMIESCYSSES-----WVCVPASTYCNNA 377
QY 326 VEGPYAFAGRGVYDIRHPYDDP-----TPPSYNYKFLAKDSVMDAIGVNNIN-YTQSNND 379
Db 378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYSDYLNKPEVTEAVGAEVNGYDSCNFD 436
QY 380 VYFAFQGTGFVWP--NFIEDLEILALPVRSIY-GDADYICNWFEGGQAVSLAANSQ 436
Db 437 INRNFLFHGDMKPYHRLVPLGLE--QIPV---LIYAGDADFICNWLGNKAWTEALEWPG 491
QY 437 AAFRSAGYTPLVK-----NGVEYGETREYGNFSFTRVYEGAGHEVPYQPIASLQLENRT 491
Db 492 QAEYASAELEDLVVDNEHTGKKGIGQVKGSHGNTFMELYGGHGMVPMQDQESSLEFFNRW 551
QY 492 IFG-W 495
Db 552 LGGEW 556

```

## RESULT 7

```

US-08-967-149-2
; Sequence 2, Application US/08967149
; Patent No. 5939305
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5939305o No. 5939305disk of No. 5939305th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,149
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,452
; FILING DATE: 28-FEB-1996
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
US-08-967-149-2

```

```

Query Match 21.8%; Score 630; DB 2; Length 557;
Best Local Similarity 31.8%; Pred. No. 4.3e-49;
Matches 154; Conservative 81; Mismatches 174; Indels 76; Gaps 17;

```

```

QY 27 ANNTYRKPEKAGVCEITFGVKSYSYGVYDISP-ESHFFWFPEARNPETAPITLWLN 85
Db 132 AYDLRVKTKDPSLGI---DEGVKQYTGILDNDNKKHLFYWFPERNDPDPVWLNL 188
QY 86 GPGCSDSLIGLFEELGPGCHVNSTFDDYINPHSWNEVSNLLFLSOPLVGVFSYSDTVDGSI 145
Db 189 GPGCCSLTGLFMELGPFSSINKKIQPVYNDYAWNSNASVIFLDQPVNVGYYSNSA----- 244
QY 146 NPVTGVVENSFAGVQGRYPIDATLDTTNIAAEAAWEILQGLSLPSLDSRVQSKDF 205
Db 245 -----VSDIVAAGKDVYALLTLFFKOPF-----EYAKQDF 274
QY 206 SLWTESYGHYGFAPFHFYEQNRIANGSVNGVQLNFNSLGIINGIIDEAIAQIAPYPEF 265
Db 275 HIAGESYAGHYPIPVFASEILSHKKR-----NINLQSVLIGNGLIDGVTOYEYRPM 325
QY 266 AVNNTYGIKAVNETVYNYMKFANPONGCDLISTCKQTNRTALADYALCAEATNMCRDN 325
Db 326 ACGDGGYPAVLDESSCQSD--NALPR-COSMIESCYSSES-----WVCVPASTYCNNA 377
QY 326 VEGPYAFAGRGVYDIRHPYDDP-----TPPSYNYKFLAKDSVMDAIGVNNIN-YTQSNND 379
Db 378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYSDYLNKPEVTEAVGAEVNGYDSCNFD 436
QY 380 VYFAFQGTGFVWP--NFIEDLEILALPVRSIY-GDADYICNWFEGGQAVSLAANSQ 436
Db 437 INRNFLFHGDMKPYHRLVPLGLE--QIPV---LIYAGDADFICNWLGNKAWTEALEWPG 491
QY 437 AAFRSAGYTPLVK-----NGVEYGETREYGNFSFTRVYEGAGHEVPYQPIASLQLENRT 491
Db 492 QAEYASAELEDLVVDNEHTGKKGIGQVKGSHGNTFMELYGGHGMVPMQDQESSLEFFNRW 551
QY 492 IFG-W 495
Db 552 LGGEW 556

```

## RESULT 8

```

US-08-309-341-4
; Sequence 4, Application US/08309341
; Patent No. 5594119
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5594119o No. 5594119disk of No. 5594119th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,341
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435

```

ATTORNEY/AGENT INFORMATION:  
NAME: Lowney, Karen A.  
REGISTRATION NUMBER: 31,274  
REFERENCE/DOCKET NUMBER: 4247,000-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 557 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Aspergillus Niger  
US-08-309-341-4

Query Match 21.6%; Score 623; DB 1; Length 557;  
Best Local Similarity 32.0%; Pred. No. 1.9e-48;  
Matches 155; Conservative 79; Mismatches 175; Indels 76; Gaps 18;

QY 27 ANNVTRYKEPGAEGCETTPGVKSYSGYVDTSP-ESHTFFEFARHNPETAPITLWLN 85  
Db 132 AYDLRYKTDPSLGI---DPGVKQYTGILDDNENDKHLFYWFESRNDPENVVLWLN 188  
QY 86 GPGSDSLIGLFEELGCHVNSTFDDYINPHSWNEVSNLLFSOPLGVGFYSYSDIVDGS 145  
Db 189 GPGCSLSLIGLFEELGCHVNSTFDDYINPHSWNEVSNLLFSOPLGVGFYSYSDIVDGS 145  
QY 146 NPVTGVVENSFAGVQGRYPTIDATLIDTNTLAAEAWEILQGLSLGLSLDSRVQSKDF 205  
Db 245 -----VSDTVAAGKDVYALLTLFFKQFP-----EYAKQDF 274  
QY 206 SLWTESYGGHYGPAFFNHFYEQNERIANGSVNGVQLNFNSLGIINGIIDEAIAQYYPEF 265  
Db 275 HIAGESYAGHYIPVFASEILSHKRR-----NINLQSVLIGNGLTDLGLTQYEYRPM 325  
QY 266 AVNNTYGIKAVNEVTVNYMKFANQMPNGCDLISTCKQTNRTALADYALCAEATNMCRDN 325  
Db 326 ACDG-GYPVLDE-GSCQAMDNALPR-CQSMIESCYSSA-----WVCVPASIYCNA 377  
QY 326 VEGPYAFAGRGVYDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGVIN-YTQSNND 379  
Db 378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYVSDYLNKTEVIEAVGAEVNGYDSCNFD 436  
QY 380 VYFAFOQTGDFVWP--NFIEDLEILALPVRVSLTY-GDADYICNWFGGQAVSLAANSQ 436  
Db 437 INRNFLHGDGMKPYHRLVPGLE--QIPV---LIYAGDADFICNLGNKAWTEALEWPG 491  
QY 437 AAQFRSAGYTPLVK-----NGVEYGETREYNFSFTRYEAGHEVPYQPTASLQLFNRT 491  
Db 492 QAEYASAKLEDLVVNEHKGKKGQVKSNGHNTFMRLYGGGHVMPMDQPESSLEFFNRW 551  
QY 492 IFG-W 495  
Db 552 LGGEW 556

RESULT 9  
US-08-608-267-4  
Sequence 4, Application US/08608267  
Patent No. 5688663  
GENERAL INFORMATION:  
APPLICANT: Yaver, Debbie Sue  
APPLICANT: Thompson, Sheryl Ann  
TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5688663o No. 5688663disk of No. 5688663th America, Inc.  
STREET: 405 Lexington Avenue, Suite 6400  
CITY: New York  
STATE: New York



```
RESULT 10
US-08-608-452-4
; Sequence 4, Application US/08608452
; Patent No. 5693510
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 56935100 No. 5693510disk of No. 5693510th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,452
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
; US-08-608-452-4

Query Match 21.6%; Score 623; DB 1; Length 557;
Best Local Similarity 32.0%; Pred. No. 1.9e-48;
Matches 155; Conservative 79; Mismatches 175; Indels 76; Gaps 18;

QY 27 ANNVTIRKEPGAEGVCETTPGVKSYGVDTSP--ESHPTFFWFFFAHNPETAPITLWN 85
Db 132 AYDLRVKKTDPSSLGI--DPGVKQYGYLDNENDKHLFYEFFESRNDPNDPVVYLWN 188
QY 86 GPGGSDSLGLEELGCPCHVNSTFDDYINPHSWNEVSNLLFSLQPLGVGFYSYSDVDSI 145
Db 189 GPGGCSLTGLFMELGPSSINKKIOPVYNDYAWNSASVIFLDPQVNVGYSYNSA---244
QY 146 NPTVGTWVNSSFAGVQGYPTIDATLIDTTLNAAEANEILQGLFLGSLDSRVQSKDF 205
Db 245 -----VSDTVAAGKDVYALLTLFFKQEP-----EYAKQDF 274
QY 206 SLWTSYGGHYGAPFNFHYEQNERIANGSVNGVQLNLSGLTINGIIDEAIOAPYYPEF 265
Db 275 HIAGESAGHYIPVFASELSHKKR-----NINQSVLGNGLDGLTQYEEYRPM 325
QY 266 AVNNTYIGKAVNETVYNTWKFNQMPNGCQDLISTCKOTNRVALADYALCABATNMCRDN 325
Db 326 ACGDG-GYPAVLDE-GSCQAMDNALPR-CQSMIESCYSSESA----WVCVPASIYCNNA 377
QY 326 VEGPYVAFAGVYDIRPYDDP-----TPPSYNNKFLAKUSVMDAIGNIN-YTQSNND 379
Db 378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYVSDYLNKTEVIRAVGAENVGYDCNFD 436
```

```
QY 380 VYAFQQTGDFWVP--NFIEDLELALPVRVSLIY-GDADYICNWFQGVQAVSLAANYSQ 436
Db 437 INRNFLFHGDWMKPYHRLVPGILLE--QIPV---LIYAGDAFICNWLGNKAWTEALEWPG 491
QY 437 AAQFRSAGYTPKLV-----NGVEYGETREYGNFSEFTRYVEAGHEVPPYQPIASLQLFNRT 491
Db 492 QAEYASAKLEDLVVNEHKGKIGQVKSNGNFTMRDLYGGGHMVPMDDQPESSLEFFNRW 551
QY 492 IFG-W 495
Db 552 LGGEW 556
```

```
RESULT 11
US-08-608-224-4
; Sequence 4, Application US/08608224
; Patent No. 5705376
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 57053760 No. 5705376disk of No. 5705376th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,224
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
; US-08-608-224-4
```

```
Query Match 21.6%; Score 623; DB 1; Length 557;
Best Local Similarity 32.0%; Pred. No. 1.9e-48;
Matches 155; Conservative 79; Mismatches 175; Indels 76; Gaps 18;

QY 27 ANNVTIRKEPGAEGVCETTPGVKSYGVDTSP--ESHPTFFWFFFAHNPETAPITLWN 85
Db 132 AYDLRVKKTDPSSLGI--DPGVKQYGYLDNENDKHLFYEFFESRNDPNDPVVYLWN 188
QY 86 GPGGSDSLGLEELGCPCHVNSTFDDYINPHSWNEVSNLLFSLQPLGVGFYSYSDVDSI 145
Db 189 GPGGCSLTGLFMELGPSSINKKIOPVYNDYAWNSASVIFLDPQVNVGYSYNSA---244
QY 146 NPTVGTWVNSSFAGVQGYPTIDATLIDTTLNAAEANEILQGLFLGSLDSRVQSKDF 205
```

ORIGINAL SOURCE:  
ORGANISM: Aspergillus Niger  
US-08-967-149-4

Query Match 21.6%; Score 623; DB 2; Length 557;  
Best Local Similarity 32.0%; Pred. No. 1.9e-48;  
Matches 155; Conservative 79; Mismatches 175; Indels 76; Gaps 18;

27 ANNVTIRYKPPGAGVCEETTPGVKSYSGYVDTSP--ESHTFTWFEARHNPETAPITWLN 85  
132 AYDLRVKKTDPSSLGI--DPGVKQYTGYYLDDNENKHLFWFEPFERNDEPVPVWLN 188  
86 GPGCSLSLGLFEELGCPCHVNSTFDDYINPHSNVEYNLLFLSOPLGFGVSYSDTVGSI 145  
189 GPGCSLSLGLFEELGCPCHVNSTFDDYINPHSNVEYNLLFLSOPLGFGVSYSDTVGSI 244  
146 NPVTGVVENSFFAGVQGRYPTIDATLIDTFTNLAEEAWEILOGLFSLGSLDSRVOSKDF 205  
245 -----VSDTVAAGKDVYALLTLFFKQFP----EYAKQDF 274  
206 SLMTESYGGHYGPAFNHFYEQNERIANGSVNGVQLNFSNLGTLINGLIIDAIQAPYYPF 265  
275 HIAGESYAGHYIPVFASEILSHKKR-----NINLQSVLIGNGLTDLGTQYTYRPM 325  
266 AVNNTYGLKAVNETVYNYMKFANOMPNGCODLISTCKOTNKTALADYALCAEATNMCRDN 325  
326 ACGDG-GYPAVLDE-GSCQAMDNALPR-COSMIESCYSSESA-----WVCVPASLYCNA 377  
326 VEGFYAFAGRGVYDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGVNNIN-YTQSNND 379  
378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYVSDYLNTKTEVIEAVGAEVNGYDSCNFD 436  
380 VYAFQOTGDFVWP--NFIEDLEILALPVRVSLIY-GDADYICNWFGGQAVSLAANYSQ 436  
437 INRNFLPHGDWMPYHRLVPGLE--QIPV--LIYAGDADFICNWLGNKAWTEALEWPG 491  
437 AAQFRSAGYTPLVK-----NGVEYGETREYGNFSTRVYAGHEVPYQPIASLQLFNRT 491  
492 QAEYASAKLEDLVVVENEHKKGKIGQVKSNGFTMRLYGGGHMVPMDQPESSLEFENRW 551  
492 IFG-W 495  
552 LGGEW 556

RESULT 13  
US-09-640-305-4  
Sequence 4, Application US/09640305  
Patent No. RE37447  
GENERAL INFORMATION:  
APPLICANT: Fleer, Reinhard  
Fournier, Alain  
Yeh, Patrice  
TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR PREPARATION AND USE  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Rd. 3c43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19002  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/640,305  
FILING DATE: 16-Aug-2000  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,673

245 -----VSDTVAAGKDVYALLTLFFKQFP----EYAKQDF 274  
206 SLMTESYGGHYGPAFNHFYEQNERIANGSVNGVQLNFSNLGTLINGLIIDAIQAPYYPF 265  
275 HIAGESYAGHYIPVFASEILSHKKR-----NINLQSVLIGNGLTDLGTQYTYRPM 325  
266 AVNNTYGLKAVNETVYNYMKFANOMPNGCODLISTCKOTNKTALADYALCAEATNMCRDN 325  
326 ACGDG-GYPAVLDE-GSCQAMDNALPR-COSMIESCYSSESA-----WVCVPASLYCNA 377  
326 VEGFYAFAGRGVYDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGVNNIN-YTQSNND 379  
378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYVSDYLNTKTEVIEAVGAEVNGYDSCNFD 436  
380 VYAFQOTGDFVWP--NFIEDLEILALPVRVSLIY-GDADYICNWFGGQAVSLAANYSQ 436  
437 INRNFLPHGDWMPYHRLVPGLE--QIPV--LIYAGDADFICNWLGNKAWTEALEWPG 491  
437 AAQFRSAGYTPLVK-----NGVEYGETREYGNFSTRVYAGHEVPYQPIASLQLFNRT 491  
492 QAEYASAKLEDLVVVENEHKKGKIGQVKSNGFTMRLYGGGHMVPMDQPESSLEFENRW 551  
492 IFG-W 495  
552 LGGEW 556

RESULT 12  
US-08-967-149-4  
Sequence 4, Application US/08967149  
Patent No. 5939305  
GENERAL INFORMATION:  
APPLICANT: Yaver, Debbie Sue  
Thompson, Sheryl Ann  
TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 59393050 No. 59393050disk of No. 59393050th America, Inc.  
STREET: 405 Lexington Avenue, Suite 6400  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967,149  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,452  
FILING DATE: 28-FEB-1996  
APPLICATION NUMBER: US 08/309,341  
FILING DATE: 20-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lowney, Karen A.  
REGISTRATION NUMBER: 31,274  
REFERENCE/DOCKET NUMBER: 4247.000-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 557 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein

```

; FILING DATE: 06-FEB-1995
; APPLICATION NUMBER: WO PCT/FR93/00623
; FILING DATE: 23-JUN-1993
; APPLICATION NUMBER: FR 92/07785
; FILING DATE: 25-JUN-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME: Smith, Julie K.
;   REGISTRATION NUMBER: 38,619
;   REFERENCE/DOCKET NUMBER: ST92040-US
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (610)454-3839
;   TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 491 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-640-305-4

Query Match          19.7%; Score 568.5; DB 1; Length 491;
Best Local Similarity 31.1%; Pred. No. 1.5e-43;
Matches 151; Conservative 66; Mismatches 184; Indels 85; Gaps 14;

QY 27 ANNVTRYKPPGAGVCEITPGVKSYSYVDTSPESTFFWFTEARHNPETAPITWLNG 86
Db 62 AYSLRIKPLDPKSLGV----DTVKWSSGYLDYQDSKHFYWFESRNDPNDPVILNG 117
QY 87 GPQSDSLIGLFEELGPHVNSTFDDYINPHSNWNSNLLFSLQPLGVGFYSYDVTGDSIN 146
Db 118 GPCSSFVGLFFELGPGSSIGADLKPIYNPYSNWSNASVIFLDQPVGVGFYSYGD----- 170
QY 147 PVTGVVENSFAGVQGRYPTIDATLIDTNLAEEAAWEILQGLSLGLPSLDSRVQSKDFS 206
Db 171 -----SKVSTDDAAKDVYIFLDLFFERPHL-----RNNDHF 203
QY 207 LWTESYGGHYGPAFFNHFEQNERIANGSVNGVQLNFNSLGIINGIDEAIOAPYYPEFA 266
Db 204 ISGESYAGHYLPKIAH-----EIAVVAEDSSFNLSVVLGNGFTDPLTQYIYEPMA 256
QY 267 VNNYGIKAVNETVYNYMFKFANOMPNGCOD-----LISTCKOTNTALADYALCA 316
Db 257 CGEG-GYPAVLE-----PEDCLDMNENLPLCLSLVDRCYKSH-----SVFSCV 298
QY 317 EATNMCRDNVGPPYAFAGGVYDIR---HPYDDP---TPPSYNNKFLAKDSVMDAIGV 369
Db 299 LADRYCEQQTIG-VYEKSGRNPYDIRSKCEAEDDSGACYQEETIISDYLNQEEVQRAIGT 357
QY 370 NINYTOS--NDVYAFQQTGDFVWPNFIEDLEEILALPVRSVLIYGDADYICNWFQGVAV 428
Db 358 DVSSFGCCSSDVIGIGFAFTGDGSP--FHQYVAELLDQDINVLIIYAGDKYICNWLGNLAW 416
QY 429 SLAANYSQAAQFRSAGYTPLKVGVE--YGETREYGNFSTRVRYEAGHEVPYQPIASLQ 486
Db 417 TEKLEWRYNEEYKKQVLRTWKSEETDETIGETKSYGPLYLRIYDAGHMVPHDPQNSLQ 476
QY 487 LENRTI 492
Db 477 MYNSWI 482

```

## RESULT 14

```

US-08-360-673-4
; Sequence 4, Application US/08360673
; Patent No. 5679544
; GENERAL INFORMATION:
;   APPLICANT: Fleer, Reinhard
;   APPLICANT: Fournier, Alain
;   APPLICANT: Yeh, Patrice
;   TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
;   PREPARATION AND USE
;   NUMBER OF SEQUENCES: 17

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,673
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR93/00623
; FILING DATE: 23-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/07785
; FILING DATE: 25-JUN-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME: Smith, Julie K.
;   REGISTRATION NUMBER: 38,619
;   REFERENCE/DOCKET NUMBER: ST92040-US
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (610)454-3839
;   TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 491 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-360-673-4

```

```

Query Match          19.7%; Score 568.5; DB 1; Length 491;
Best Local Similarity 31.1%; Pred. No. 1.5e-43;
Matches 151; Conservative 66; Mismatches 184; Indels 85; Gaps 14;

```

```

QY 27 ANNVTRYKPPGAGVCEITPGVKSYSYVDTSPESTFFWFTEARHNPETAPITWLNG 86
Db 62 AYSLRIKPLDPKSLGV----DTVKWSSGYLDYQDSKHFYWFESRNDPNDPVILNG 117
QY 87 GPQSDSLIGLFEELGPHVNSTFDDYINPHSNWNSNLLFSLQPLGVGFYSYDVTGDSIN 146
Db 118 GPCSSFVGLFFELGPGSSIGADLKPIYNPYSNWSNASVIFLDQPVGVGFYSYGD----- 170
QY 147 PVTGVVENSFAGVQGRYPTIDATLIDTNLAEEAAWEILQGLSLGLPSLDSRVQSKDFS 206
Db 171 -----SKVSTDDAAKDVYIFLDLFFERPHL-----RNNDHF 203
QY 207 LWTESYGGHYGPAFFNHFEQNERIANGSVNGVQLNFNSLGIINGIDEAIOAPYYPEFA 266
Db 204 ISGESYAGHYLPKIAH-----EIAVVAEDSSFNLSVVLGNGFTDPLTQYIYEPMA 256
QY 267 VNNYGIKAVNETVYNYMFKFANOMPNGCOD-----LISTCKOTNTALADYALCA 316
Db 257 CGEG-GYPAVLE-----PEDCLDMNENLPLCLSLVDRCYKSH-----SVFSCV 298
QY 317 EATNMCRDNVGPPYAFAGGVYDIR---HPYDDP---TPPSYNNKFLAKDSVMDAIGV 369
Db 299 LADRYCEQQTIG-VYEKSGRNPYDIRSKCEAEDDSGACYQEETIISDYLNQEEVQRAIGT 357
QY 370 NINYTOS--NDVYAFQQTGDFVWPNFIEDLEEILALPVRSVLIYGDADYICNWFQGVAV 428
Db 358 DVSSFGCCSSDVIGIGFAFTGDGSP--FHQYVAELLDQDINVLIIYAGDKYICNWLGNLAW 416
QY 429 SLAANYSQAAQFRSAGYTPLKVGVE--YGETREYGNFSTRVRYEAGHEVPYQPIASLQ 486
Db 417 TEKLEWRYNEEYKKQVLRTWKSEETDETIGETKSYGPLYLRIYDAGHMVPHDPQNSLQ 476

```





TELEPHONE: 212-867-0123  
 TELEFAX: 212-878-9655  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 554 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FRAGMENT TYPE: internal  
 US-08-943-714-2

Alignment Scores:  
 Pred. No.: 1,296-268 Length: 554  
 Score: 2878.00 Matches: 536  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.10% Indels: 0  
 DB: 3 Gaps: 0

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-08-943-714-2 (1-554)

```

QY 1 CTTCCAGGAGTACACCGCGCTCGGTGAGACAGTACCCCAAGAACCCACCGG 60
DB 19 LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly 38
QY 61 GTCAGAGCTCTACACCCCAACATGTCACCATCCGGTACAGGAACCCGGGAG 120
DB 39 ValLysThrLeuThrThrAlaAsnAsnValThrLeuArgTyrLysGluProGlyAlaGlu 58
QY 121 GCGCTCGGAGACTACCCCGGTGCAAACTCTACTCTGGATATGTCGACACTCTCC 180
DB 59 GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 78
QY 181 GAGTCCCATACCTCTCTCGTTCTTCGAGCCAGACATACCCAGAACTGCACCTATC 240
DB 79 GluSerHisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIle 98
QY 241 ACATTTGGTTGAATCGTGGCCCTGGAAGGATCTTTGATCGGTCTCTCGAAGATTG 300
DB 99 ThrLeuTrpLeuAsnGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeu 118
QY 301 GGCCTTGCCATGCAATTCGACTTTGATGACTATCATCAACCTCCTCGTGGAGAGAG 360
DB 119 GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTyrAsnGlu 138
QY 361 GTCTCCAATTTACTATTCCTGTCCTCCAGCCCATTTGGAGTCGGCTTTTCATATAGTATACG 420
DB 139 ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 158
QY 421 GTTCATGGTCCATTACCTGTAACCTGGGCTGCGAAATTCGAGCTTTCGAGGAGTT 480
DB 159 ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal 178
QY 481 CAGGCGCGTACCAACCATTTGATGCTGCTGATGATACCTACCAATCTTGGCCGAGAG 540
DB 179 GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGlu 198
QY 541 GCGCTGGAGATCTCGAAGATCTCTTAGTGGAGTACCTAGCTGGAGTCTAGGGIG 600
DB 199 AlaAlaTrpGlnIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal 218
QY 601 CAGTCTAAGACTTCACTCTATGACGAGAGCTATGAGGCGCACTATGTCCTGCAATTC 660
DB 219 GlnSerLysAspPheSerLeuThrThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 238
QY 661 TTCATCATTTTACGACGAGATGAGAGAATGCCAACGCTACTGTATTAATGTTTCAG 720
DB 239 PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 258
QY 721 CTTAATTTCACTCTCTGGGAATTTAATACGGCATTCAGCAGGCGCATCCAGCCCT 780

```

```

DB 259 LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAsnGlyValAlaIleGlnAlaPro 278
QY 781 TACTACCCCTGAATTCGCTGTGAACATACCTACGGTATCAAGGCTGTCAACGAGACGTC 840
DB 279 TyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleIleValAsnGluThrVal 298
QY 841 TACAACATCATCAAGTTTCCCAACAAATGCCAAATGGTTCGAGGATTGATTCCACC 900
DB 299 TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr 318
QY 901 TCGAACAACACAAACCGCACCGCATTAGCTAGCTACGCCCTCTGCGCCGGAAGCACCAAC 960
DB 319 CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn 338
QY 961 ATGTGACAGGACAATGTGAGGGCCATACCTAGCCCTTTGCTGCTGCTGTGTGTATGAT 1020
DB 339 MetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAsp 358
QY 1021 ATTGCGCATCCATATGATGACCGGACTCCGCGCAAGTTATTACAAATTTCTGGCAAG 1080
DB 359 IleArgHisProTyrAspAspProThrProSerTyrTyrAsnLysPheLeuAlaLys 378
QY 1081 GACTCTGTGATGACGCTATCGGCGTCAACATCACTACACCCAGTCCCAATATGACGTC 1140
DB 379 AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspVal 398
QY 1141 TACTAGCTTTCACGCAACACAGCGACTTTGCTGCGCCCAACTCATCTATGCGACCGCTCGAG 1200
DB 399 TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu 418
QY 1201 GAGATCCTTCTCCCGCTGCGTCTCCTCATCTATGCGACCGCGATTCATCTGC 1260
DB 419 GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys 438
QY 1261 AACTGTTGCGCGTACAGCGCTTTCCTCGTGGAACTACTCCCAAGCGCCGCTTC 1320
DB 439 AsnTrpPheGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPhe 458
QY 1321 CGAAGCGCAGGTACACGCGCTGAAAGTCAACGCGCTCGAGTATGGGAACTCGCGAG 1380
DB 459 ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyThrArgGlu 478
QY 1381 TATGGTAATTTCTCTTCACCTCGCTATGAGCGACGCGCATGAAGTCCCATCTACACAG 1440
DB 479 TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln 498
QY 1441 CCAATCGCTCCCTGCAATTTTAAACGGCACTATCTTCGGTGGGATATCGCAGAGGC 1500
DB 499 ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTyrAspIleAlaGluGly 518
QY 1501 CAGAAGAAGATCTCGCCAGCTACAGACGATCGAGCGGTACAGCTACGCTACGATACAG 1560
DB 519 GlnLysLysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGln 538
QY 1561 TCGTCCGTCGCGTCCCTACGCTACGCGATGTCAGTGTGTGTATG 1608
DB 539 SerSerValProLeuProThrAlaThrSerMetSerSerValGlyMet 554

```

## RESULT 2

US-08-943-714-9  
 ; Sequence 9, Application US/08943714  
 ; Patent No. 6187578  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blinkovsky, Alexander  
 ; APPLICANT: Berk, Randy  
 ; APPLICANT: Rey, Michael  
 ; APPLICANT: Golightly, Elizabeth  
 ; APPLICANT: Klotz, Alan  
 ; APPLICANT: Mathisen, Thomas Erik  
 ; APPLICANT: Dambmann, Claus  
 ; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids  
 ; TITLE OF INVENTION: Encoding Same  
 ; NUMBER OF SEQUENCES: 12







```

US-08-608-267-2
; Sequence 2, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5688663o No. 5688663disk of No. 5688663th America, Inc
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,267
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowmney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
US-08-608-267-2

Alignment Scores:
Pred. No.:          6,13e-52      Length:       557
Score:             630.00         Matches:       154
Percent Similarity: 48.45%        Conservative:   81
Best Local Similarity: 31.75%     Mismatches:    174
Query Match:       21.69%         Indels:        76
DB:                1              Gaps:          17

US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-608-267-2 (1-557)
QY      79  GCACAACTGTCACCATCCGGTCAAGGAACCGGGGCAGAGGGCGTGCGAGACTACC 138
Db      132  AlalyrAspLeuArgVallyLysThrAspProglySerLeuGlyIle-----Asp 148
QY      139  CGGGGTCAAACTCCTACTCGGATATGTCGACACCTCTGCC---GAGTCCCATAACCTC 195
Db      149  ProGLyVallyLysGlnIyrThrGlyTyLeuAspAsnGluAsnAspLysHisLeuPhe 168
QY      196  TTCGTGTTCTTCGAAGCGACATFAACCCAGAACTGCACCTATCACATCTGTGGTTGAAT 255
Db      169  TyrTrpPheGluSerArgAsnAspProGluAsnAspProvalValLeuTrpLeuAsn 188
QY      256  GTGGGCCCTGGAGAGCATICTTTGATCGGTCTCTTCGAAGAGTGTGGCCCTTGGCATGTC 315
Db      189  GlyGlyProglyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 208
QY      316  AATTTCGACITTTGATGACTACATCAACCCCTCACTCGTGGGAACGAGGTCTCCAATTACIA 375
Db      316  AATTTCGACITTTGATGACTACATCAACCCCTCACTCGTGGGAACGAGGTCTCCAATTACIA 375

```

QY 1414 GCAGGCCATGAGTCCATACCTACAGCCATCGCCTCCGTAATGTTTACCGGACT 1473  
 Db 532 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPhePheAsnArgTirp 551  
 QY 1474 ATCTCGGT---TGG 1485  
 Db 552 LeuGlyGlyGluTirp 556  
 RESULT 5  
 US-08-608-452-2  
 ; Sequence 2, Application US/08608452  
 ; Patent No. 5693510  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yaver, Debbie Sue  
 ; APPLICANT: Thompson, Sheryl Ann  
 ; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 5693510 No. 5693510disk of No. 5693510th America, Inc.  
 ; STREET: 405 Lexington Avenue, Suite 6400  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10174-6401  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/608,452  
 ; FILING DATE: 28-FEB-1996  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/309,341  
 ; FILING DATE: 20-SEP-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lowney, Karen A.  
 ; REGISTRATION NUMBER: 31,274  
 ; REFERENCE/DOCKET NUMBER: 4247,000-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212 867 0123  
 ; TELEFAX: 212 867 0298  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 557 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Aspergillus Niger  
 ; US-08-608-452-2  
 Alignment Scores:  
 Pred. No.: 6,13e-52  
 Score: 630.00  
 Percent Similarity: 48.45%  
 Best Local Similarity: 31.75%  
 Query Match: 21.69%  
 DB: 1  
 US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-08-608-452-2 (1-557)  
 QY 79 GCAACAAATGTCACATCCGTCACAGAACCCGGCGAGAGCGGCTCGGAGACTACC 138  
 Db 132 AlaTyrAspLeuArgValLysLysThrAspProGlySerLeuGlyLeu-----Asp 148  
 QY 139 CCGGTGTCAATCTACTCTGATATGTCGACACCTCTCCC---GAGTCCCATACCTTC 195  
 Db 149 ProGlyValLysGlnTyrThrGlyTyrLeuAspAspAsnGluAsnAspLysHisLeuPhe 168

QY 196 TTCTGGTTCTTGAAGCCAGACATACCCAGAAACTGCACATATCATCTGTTGTTGAAT 255  
 Db 169 TyrTirpPheGluSerArgAsnAspProGluAsnAspProValValLeuTirpLeuAsn 188  
 QY 256 GGTGGCCCTGGAACCGATTCTTTCATCGGTCTCTTGAAGAGTGTGGCCCTTGCATGTC 315  
 Db 189 GlyGlyProGlyCysSerSerLeuThrGlyLeuMetGluLeuGlyProSerSerLeu 208  
 QY 316 AATTCGACTTTTGATGACTACATCAACCCCTCACTCGTGGGAACGAGGTCTCCAAATTACTA 375  
 Db 209 AsnLysLysIleGlnProValTyrAsnAspTyrAlaIlePasnSerAsnAlaSerValIle 228  
 QY 376 TTCTGTGTCGCCAGCCATTGGGAGTGGCTTTTCATATAGTATAGTACGTTGATGGTCCATT 435  
 Db 229 PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla----- 244  
 QY 436 AACCTGTAACTGGGGTCTCGAAAATTTCGAGGTTCGAGGAGTTTCAGGGCGGTACCCA 495  
 Db 244 ----- 244  
 QY 496 ACCATTGATGCCACTCTGATCGATACIACCAATCTTGGCGAGAGCCGCTTGGAGATC 555  
 Db 245 -----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu 258  
 QY 556 CTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGTGTCAGTCTAAGGACTTC 615  
 Db 259 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaLysGlnAspPhe 274  
 QY 616 ACTCTATGAGCGGAGAGTATGAGGGGACTATGGTCTGCATCTTCAATCATTTTAC 675  
 Db 275 HisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeu 294  
 QY 676 GAGCAGAAATGAGAGATTGCCAAGGTAGTGTAAATGTGTTCAGCTTAATTTCAACTCT 735  
 Db 295 SerHisLysLysArg-----AsnIleAsnLeuGlnSer 305  
 QY 736 CTGGGAATTAATTAACGGCATATPCAGAGGGGATCCAGGCCCTTACTACCTGATTC 795  
 Db 306 ValLeuIleGlyAsnGlyLeuThrAspGlyTyrThrGlnTyrGluTyrArgProMet 325  
 QY 796 GCTGTGAACAATACCTACGTATCAAGGTGTCAAGGAGACCTCTACAACTACGAAG 855  
 Db 326 AlaCysGlyAspGlyGlyTyrProAlaValLeuAspGluSerSerCysGlnSerMetAsp 345  
 QY 856 TTTCGCCAACCAATGCCAATGTTCGACGAGTTGATTTCCACCTGCAACAGACAAAC 915  
 Db 346 -----AsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerGlu 362  
 QY 916 CGCACCCGATTAGCTAGCTACGCCCTCTGCGCCGAGACCCACCATCTGACGGACAAT 975  
 Db 363 SerAla-----TipValCysValProAlaSerIleTyrCysAsnAla 377  
 QY 976 GTTCAGGGGCAATACCTACGCCCTTGTGCTGTGTGTATGATATTCGSCATCCATAT 1035  
 Db 378 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArgGlyLysCys 396  
 QY 1036 GATCACCCG-----ACTCCGCCAAAGTTATTACAAATTTCTGGCAAG 1080  
 Db 397 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 416  
 QY 1081 GACTCTGTATGAGCGGTATCGGGCTCAACATCAAC---TACACCCAGTCCCAATAATGAC 1137  
 Db 417 ProGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp 436  
 QY 1138 GTCTACTACGCTTTCACCAACAGCGGACTTTGTCTGGCC-----AATCTCATCGAA 1191  
 Db 437 IleAsnArgAsnPheLeuPheHisGlyAspTyrMetLysProTyrHisArgLeuValPro 456  
 QY 1192 GACCTCGAGGAGATCCTTGTCTCTCCCGGTGGTGTCTCCCTCATCTAT---GGCAGCC 1248  
 Db 457 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAla 471  
 QY 1249 GATTACATCTGCAACTGGTTCGGCGGTTCAGGGCGGTTTCCTCGCTCGCAACTACTCCCA 1308

```

Db      472 AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpProGly 491
QY      1309 GCGCCCGAGTCCGAAAGCGAGGTACAGCCGCTGAAAGTC-----AAC 1353
Db      492 GluAlaGluTrpAlaSerAlaGluLeuGluAspLeuValIleValAspAsnGluHisThr 511
QY      1354 GCGGTGAGTATGGGAAACTCGGAGTATGGTAATTCCTTCCTCACTCGCGTCTATGAG 1413
Db      512 GlysLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTrpGly 531
QY      1414 GCAGGCATGAAGTCCCATCTACCGCCATCCGCTCCCTCGCAATTTGTTTAAACCGGACT 1473
Db      532 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPhePheAsnArgTirp 551
QY      1474 ATCTTCGGT---TGG 1485
Db      552 LeuGlyGlyGluTrp 556

```

## RESULT 6

```

US-08-608-224-2
; Sequence 2, Application US/08608224
; Patent No. 5705376
; GENERAL INFORMATION:
; APPLICANT: Iaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5705376 No. 5705376disk of No. 5705376th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,224
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
; US-08-608-224-2

```

```

Alignment Scores:
Pred. No.: 6,13e-52 Length: 557
Score: 630.00 Matches: 154
Percent Similarity: 48.45% Conservative: 81
Best Local Similarity: 31.75% Mismatches: 174
Query Match: 21.69% Indels: 76
DB: 1 Gaps: 17

```

```

US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-608-224-2 (1-557)
QY      79 GCAACAATGTCACCATCCGGTACAAAGAACCCGGGCGAGAGGGCTTCGAGACTACC 138
Db      132 AlaTyrAspLeuArgValLysLysThrAspProGlySerLeuGlyLe-----Asp 148
QY      139 CCGGGTGTCAAACTCTACTCTGGATATGTCGACACCTCTCC---GAGTCCCATACCTTC 195
Db      149 ProGlyValLysGlnTyrThrGlyTyrLeuAspAsnGluAsnAspLysHisLeuPhe 168
QY      196 TTCCTGGTCTTCGAAGCAGACATAACCCACAACACTGCACCTATACACATCTGGTTGAAT 255
Db      169 TyrTrpPheGluSerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsn 188
QY      256 GGTGGCCCTCGAAGCGATTCTTTGATCGGTCTCTTCGAAGAGTGGCGCCCTTCGCATGTC 315
Db      189 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 208
QY      316 AATTCGACTTTTGATGACTACATCAACCTCTACCTCGTGGAAACGAGGCTCTCAATTTACTA 375
Db      209 AsnLysLysIleGlnProValTyrAsnAspTyrAlaTrpAsnSerAsnAlaSerValIle 228
QY      376 TTCCTGTCCCGAGCCATTGGGAGTCGGCTTTTCATATAGTAGTATACGTTGATGGTCCATT 435
Db      229 PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla----- 244
QY      436 AACCTCTAACTGGGTCGTCGAAATTCGAGCTTTCAGGAGTTCAGGGCGCGGTACCCA 495
Db      244 ----- 244
QY      496 ACCATTGATGCCACTCTGATCGATACWACCANCTTCGCGCAGAGGCGGTGGAGATC 555
Db      245 -----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu 258
QY      556 CTGCAAGGATTCTCTAGTGGACTACCTAGCTTGGACTCTAGGCTTGGGTGCGAGTCTAAGGACTTC 615
Db      259 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaLysGlnAspPhe 274
QY      616 AGTCTATGGACGAGAGCTATGGAGGCGACTATGGCTCTGCATCTTCAATCATTTTTCAC 675
Db      275 HisIleAlaGlyGluSerTyrAlaGlyHisIleProValPheAlaSerGluIleLeu 294
QY      676 GACGAGAATCAGAGAATGCCAACGGTAGTGTAAATGGTTCAGCTTAATTTCAACTCT 735
Db      295 SerHisLysLysArg-----AsnIleAsnLeuGlnSer 305
QY      736 CTGGGAATTATTAAACGCATCATCGAGAGCGCATCCAGCCCTTACTTACTACCTGATTC 795
Db      306 ValLeuIleGlyAsnGlyLeuThrAspGlyTyrThrGlnTyrGluTyrTyrArgPromet 325
QY      796 GCTGTGAACAATATACCTACGGTATCAAGGCTGTCAAGCAGAGACCGCTCTACAACACTACATCAAG 855
Db      326 AlaCysGlyAspGlyGlyTyrProAlaValLeuAspLysSerSerCysGlnSerMetAsp 345
QY      856 TTTCGCCAACCAATGCCAATGTTGCCAGGATTTGATTTCCACCTGCAACAGACAAC 915
Db      346 -----AsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerGlu 362
QY      916 CGCACCAGCATTAGCTGACTAGCGCCCTTCGCGGAAGCCACCAACATGTCAGGGACAT 975
Db      363 SerAla-----TrpValCysValProAlaSerIleTyrCysAsnAla 377
QY      976 GTTAGGGGCCCATACCTACGCCCTTTGCTGCTGCTGGTGTGTATGATATTCGGCATCATAT 1035
Db      378 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValargGlyLysCys 396
QY      1036 GATGACCCG-----ACTCCGCAAGTTATTACAACAATAATTTCTGGCAAG 1080
Db      397 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 416
QY      1081 GACTCTGTGATGAGCTATCGCGCTCAACATCAAC---TACACCAGTCCCATATATGAC 1137
Db      417 ProGluValIleGluAlaValGlyAlaGluAlaAsnGlyTyrAspSerCysAsnPheAsp 436

```

```

QY      1138  GHCTACTAGCGCTTTCAGCAAAACAGCGCGACTTGTCTCGGCC-----AACTTCATCGAA 1191
      :::      |||      |||||:::  |||      :::
Db      437  IleAsnArgAsnPheLeuPheHisGlyAspTrpMetLysProTyrHisArgLeuValPro 456
      :::      |||      |||||:::  |||      :::
QY      1192  GACCTCGAGGAGATCCTTGCTCTCCCGGCGGTCTCCCTCATCTAT---GGCGAGCGC 1248
      |||      |||      |||||:::  |||||:::  |||||:::
Db      457  GlyLeuLeuGlu-----GluIleProVal-----LeuIleTyrAlaGlyAspAla 471
      |||      |||      |||||:::  |||||:::  |||||:::
QY      1249  GATTACATCTCGCAACTGGTTCGGCGGTTCAGGCGGTTCCTCGCTCGGAAGTACTCCCAA 1308
      |||||:::  |||||:::  |||      |||      |||      |||      |||      |||
Db      472  AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpProGly 491
      |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
QY      1309  GCGCGCCAGTTCGGAACGCGGATACACGCCCTCGAAAGTC-----AAC 1353
      |||||:::  |||||:::  |||      |||      |||      |||      |||      |||
Db      492  GluAlaGluTyrAlaSerAlaGluLeuGluAspLeuValIleValAspAsnGluHisThr 511
      |||||:::  |||||:::  |||      |||      |||      |||      |||      |||
QY      1354  GCGCTCGAGTATGGGAAACTCGCGAGTAGTATGGTAATTTCTCCTTCACCTCGCGTCTATGAG 1413
      |||||:::  |||||:::  |||      |||      |||      |||      |||      |||
Db      512  GlyLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly 531
      |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
QY      1414  GCAGGCCATGAAGTCCCATACTACACGCCCATCGCCCTCGCAATGTGTTTACCGGACT 1473
      |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
Db      532  GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPhePheAsnArgTrp 551
      |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
QY      1474  ATCTTCGGT---TGG 1485
      :::      |||      |||
Db      552  LeuGlyGlyGluTrp 556
      |||      |||

RESULT 7
US-08-967-149-2
; Sequence 2, Application US/08967149
; Patent No. 5939305
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5939305 No. 5939305disk of No. 5939305th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,149
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,452
; FILING DATE: 28-FEB-1996
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

Db 363 SerAla-----TrrValCysValProAlaSerIleTyrCysAsnAla 377
QY 976 GTTGGGGCCACTACTACGCCITTCGTGGTGGTGTGTATGATATTCGGCAICCATAT 1035
Db 378 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArgGlyLysCys 396
QY 1036 GATGACCGG-----ACTCGCGCAAGTTATTACACAAATTTCTGGCAAG 1080
Db 397 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 416
QY 1081 GACTCTGTCTCAGGAGCTATCGGCTCAACATCAC---TACACCCAGTCCCAAAATGAC 1137
Db 417 ProGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp 436
QY 1138 GTCCTACTACGCTTTCAGCAAAACAGCGACCTTTGCTGGCCG-----AATTCATCGAA 1191
Db 437 IleAsnArgAsnPheLeuPheHisGlyAspIrrpMetLysProTyrHisArgLeuValPro 456
QY 1192 GACCTCGAGGAGATCCTGTCTCCCGCTGGGTGCTCTCCCTCATCTAT---GGCGACGCC 1248
Db 457 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla 471
QY 1249 GATTACATCTCAACTGTTTCGGCGGTACAGCGGTTTCCTCGCTGCGAATCTACGCCAA 1308
Db 472 AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpProGly 491
QY 1309 GCCGCCAGTTCGGAAGCGGAGGTACACGCCCTCGAAGTC-----AAC 1353
Db 492 GlnAlaGluTyrAlaSerAlaGluLeuGluAspLeuValIleValAspAsnGluHisThr 511
QY 1354 GCGTCGAGTAGNGGGAACATPCGCGAGTATGATGTAATTTCTCTCACTCGCGTCTATGAG 1413
Db 512 GlyLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly 531
QY 1414 GCAGGCGCATGACGCCATACACTACACGCCCATCCCTCCCTCGCAATTTTAAACCGGACT 1473
Db 532 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPheAsnArgTrp 551
QY 1474 ATCTTCGGT---TGG 1485
Db 552 LeuGlyGlyGluTrp 556

```

## RESULT 8

```

US-08-309-341-4
; Sequence 4, Application US/08309341
; Patent No. 5594119
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5594119 No. 5594119 disk of No. 5594119th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,341
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123

```

```

; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORGANISM: Aspergillus Niger
; ORGANISM SOURCE:
; US-08-309-341-4

```

## Alignment Scores:

```

Pred. No.: 2,9e-51 Length: 557
Score: 623.00 Matches: 155
Percent Similarity: 48.25% Conservative: 79
Best Local Similarity: 31.96% Mismatches: 175
Query Match: 21.45% Indels: 76
DB: 1 Gaps: 18
US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-309-341-4 (1-557)

```

```

QY 79 GCAACAATGTCAACATCGGTACAGGAACCCGGGCGAGGCGGTCTGCCGAGACTACC 138
Db 132 AlaTyrAspLeuArgValLysLysThrAspProSerSerLeuGlyIle-----Asp 148
QY 139 CCGGTGTCAAAATCCTACTCTGGATATGTGCACACTCTCTCCG---GAGTCCCATACCTTC 195
Db 149 ProGlyValLysGlnTyrThrGlyTyrLeuAspAspAsnGluAsnAspLysHisLeuPhe 168
QY 196 TTCGTGTTCTTCGAGCCAGACATACCCAGAACTGCACCTATCACAATTTGTGTTGAT 255
Db 169 TyrTrpPheGluSerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsn 188
QY 256 GGTGGCCCTGGAAGCGATTCTTGTATCGTCTCTTCGGAAGAGTTGGGCCCTTGCATGTC 315
Db 189 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 208
QY 316 AATTGCGACTTTTGATGACTACATCAACCTCTACCTGCGAGAGGTCTCCAATTTACTA 375
Db 209 AsnLysLysIleGlnProValTyrAsnAspTyrAlaTrpAsnSerAsnAlaSerValIle 228
QY 376 TTCCTGTGCCAGCCATTCGGGAGTCGGCTTTTCATATAGTATAGCGGTTGATGGTTCATT 435
Db 229 PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla-----244
QY 436 AACCTGTAACTGGGGTCTGCGAAAATTTCGAGCTTCGAGAGTTCAGGGCGCGGTACCCA 495
Db 244 -----244
QY 496 ACCATTGATGCCACTCTGATCGATACCTACCAATCTGCGCAGAGCGCGTTGGAGATC 555
Db 245 -----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu 258
QY 556 CTGCAAGGATTCCTTAGIGGACTACCTAGCTTGGACTCTAGGGTGCAGTCTAAGGACTTC 615
Db 259 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaLysGluAspPhe 274
QY 616 AGTCTATGGACGAGAGCTATGGAGGCGCACTATGGTCCTGATCTTCAATCATTTTAC 675
Db 275 HisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeu 294
QY 676 GAGCAGAAATGAGAAATTCGCAACCGGTAGTGTAAATGGTTCAGCTTAAATTTCACTCT 735
Db 295 SerHisLysLysArg-----AsnIleAsnLeuGlnSer 305
QY 736 CTGGGATATTATACGGCATCATCGACGAGCGATCCAGGCCCTTACTACCTCGAATC 795
Db 306 ValLeuIleGlyAsnGlyLeuThrGlnTyrGluTyrAspGlyLeuThrGlnTyrAspGly 325
QY 796 GCTGTGAACAATACCTACGCTATCAAGGCTGTCAACGAGACCGCTCTACACTACATGAAG 855
Db 326 AlaCysGlyAspGly---GlyTyrProAlaValLeuAspGlu---GlySerCysGlnAla 343

```

```
QY 856 TTTCGACCAATGCAATGGTCCAGGATTGATTCCACCTGCAACAGACAAAC 915
Db 344 MetAspAsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerSerGlu 362
QY 916 CGCACCGCATTTAGCTGACTACGCGCTCTGCGCGAAGCCACCAACATGTGACGGACAAT 975
Db 363 SerAla-----TrpValCysValProAlaSerIleTyrCysAsnAla 377
QY 976 GTTGAGGGGCATACATACGCTTTGCTGCTGGTGGTGTATGATATTCGGCATCATAT 1035
Db 378 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArgGlyLysCys 396
QY 1036 GATGACCCG-----ACTCCGCCAAGTTATTACACAAATTTCTGGCAAG 1080
Db 397 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspIleLeuAsnLys 416
QY 1081 GACTCTGTATGACGCTATCGCGCTCAACATCAAC---TACACCCAGTCACATATGAC 1137
Db 417 ThrGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp 436
QY 1138 GTCTACTACCTTCCAGCAAAACAGCGGACTTTGCTCGGCC-----AACTTCATCGAA 1191
Db 437 IleAsnArgAsnPheLeuPheHisGlyAspTrpMetLysProTyrHisArgLeuValPro 456
QY 1192 GACTCTGAGGATGCTCTGCTCCCGCTGCTGCTCCCTCACTAT---GGCGACGCC 1248
Db 457 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla 471
QY 1249 GATTACATCTGCAACTGGTTCGGCGGTACAGCGGCTTTCCTCGCTGGCACTATCCCAA 1308
Db 472 AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpProGly 491
QY 1309 GCCGCCAGTTCGCAACCGCAGGTACACGCCCTGGAAGTC-----AAC 1353
Db 492 GlnAlaGluTyrAlaSerAlaLysLeuGluAspLeuValValGluAsnGluHisLys 511
QY 1354 GCGCTGAGTATGGGAACCTCGCGAGTATGGTAAATTTCTGCTCACTCGCGTCTATGAG 1413
Db 512 GlyLysLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly 531
QY 1414 GCAGGCCATGAATCCCATCTACACGCCATCCCTCCCTCGCAATTTGTTAACGGGACT 1473
Db 532 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPhePheAsnArgTrp 551
QY 1474 ATCTTCGGT---TGG 1485
Db 552 LeuGlyGlyGluTrp 556

RESULT 9
US-08-608-267-4
; Sequence 4, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5688663o No. 5688663disk of No. 5688663th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,267
; FILING DATE: 28-FEB-1996
```

```
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/309,341
FILING DATE: 20-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowrey, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247,000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 557 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Aspergillus Niger
US-08-608-267-4

Alignment Scores:
Pred. No.: 2,9e-51
Score: 623.00
Length: 557
Matches: 155
Percent Similarity: 48.25%
Conservative: 79
Best Local Similarity: 31.96%
Mismatch: 175
Query Match: 21.45%
Indels: 76
Gaps: 1
DB: 18

US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-608-267-4 (1-557)
QY 79 GCAACAAATGTCACATCCCGTACAGGAACCCGGGCGAGAGCGGTCTGCGAGACTACC 138
Db 132 AlaTyrAspLeuArgValLysThrAspSerSerLeuGlyIle-----Asp 148
QY 139 CCGGTGTCAATCTCTACTCTGGATATGTCACACCTCTCCC---GACTCCCATACCTTC 195
Db 149 ProGlyValLysGlnTyrThrGlyTyrLeuAspAsnGluAsnAspLysHisLeuPhe 168
QY 196 TTCTGGTCTTCGAGCCAGACATACCCAGAACTGCACCTATCACATTTGTGGTGAAT 255
Db 169 TyrTrpPheGluSerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsn 188
QY 256 GGTGCGCCCTGGAAGCGATTCTTTGATCGGTCTCTCGAAGAGTTCGGCGCTTGCATGTC 315
Db 189 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 208
QY 316 AATTCGACITTTGATGACTACATCAACCTCTACTCTGGAAGCGGTCTCCCAATTTACTA 375
Db 209 AsnLysLysIleGlnProValTyrAsnAspTyrAlaTrpAsnSerAsnAlaSerValIle 228
QY 376 TTCTGTGTCACGCCCATTTGGGATCGGCTTTTCATATAGTATGATACGCTTTCATGGTCCATT 435
Db 229 PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla-----244
QY 436 AACCTGTAACTGGGCTGCTCGAAATTCGAGCTTTGAGAGGTTCAGGCGCGGTACCCA 495
Db 244 -----244
QY 496 ACCATTGATGCCACTCTGATCATACCTACCTATCGCGCAGAGCGCGCTTGGGAGATC 555
Db 245 -----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu 258
QY 556 CTGCAAGGATTCCTTAGTGGACTACTAGCTTTGGACTCTTAGGGTGCAGCTCAAGGACTTC 615
Db 259 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaLysGlnAspPhe 274
QY 616 AGTCTATGACCGAGAGCTATGAGGCGACTATGTCCTGCTGCTTCATCATTTTAC 675
Db 275 HisIleAlaGlySerTyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeu 294
QY 676 GAGCACAATGAGAGATTCCTCCCAACGGTAGTGTATGTTGTTTCACTTCAACTCT 735
```



```
; GENERAL INFORMATION:
; APPLICANT: Iaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5705376o No. 5705376disk of No. 5705376th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,224
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
; US-08-608-224-4

Alignment Scores:
Pred. No.:          2.9e-51      Length:          557
Score:             623.00       Matches:         155
Percent Similarity: 48.25%     Conservative:    79
Best Local Similarity: 31.96%   Mismatches:     76
Query Match:        21.45%     Indels:         76
DB:                  1         Gaps:           18

US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-608-224-4 (1-557)
QY      79  GCACCAATGTCACCATTCCGGTACAAGAACCGGGGCAGAGGCGCTGCGAGACTACC 138
||| ||| :||||: :||||: :|||: :|||: :|||:
Db      132 AlATyArAspLeuArgValLysLyThrAspProSerSerLeuGLyIle-----Asp 148
||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      139  CC GG GTGTCAAATCCTACTCTGGATATGTCGACACCTCTCCC---GAGTCCCATACCTTC 195
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      149  ProGLyValLysGLnTyrThrGLyTyrLeuAspAspAsnGluAsnAspLysHisLeuPhe 168
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      196  TTC TGG GTTCTWTCGAAGCAGACATAACCCAGAACTGCACCTATCACATTGTTGGTGAAT 255
:||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db      169  TyrTrpPhePheGluSerArgAsuAspProGluAsuAspProValValLeuTripleuAsn 188
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      256  GGTGGCCCTGGAAGCGATCTCTTGATCGGTCTCTTCGAAGAGTTGGGCCCTTGCCATGTC 315
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      189  GLYGlyProGLyCySesSerSerLeuThrGLyLeuPheMetGluLeuGLyProSerSerIle 208
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      316  AATTCGACTTTTGTGATGACTACATCAACACCTCACTCGTGGAGAGGTCTCCAATTACTA 375
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      209  AshLysLysIleGLnProValTyRAsnAspTyAlaTrpAsnSerAsuAlaSerValIle 228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      376  TTCCTGTCGCCGCCCATWTGGGAGTCGGCTTTTCATATAGTGATACGGTTCATGGGTCCATT 435
```

QY	556	CTGCAAGGATTCTCTTAGTGGACTACTAGCTAGCTTGGACTCTAGGGTGCACTGCTAAGSACTTC	615
Db	259	LeuThrLeuPhePheIysGlnPhePro-----GluTyrAlaIysGluAspPhe	274
QY	616	AGTCTATGACGAGGAGGACTATGAGGAGGCACTATGCTCTGCATCTTCAATCAATTTTAC	675
Db	275	HisIleAlaGlyGlySerTyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeu	294
QY	676	GAGCAGATGAGAGAAATCCCAACGCTAGTGTATGGTGTCTACGCTTAATTTCAACTCT	735
Db	295	SerHisLysLysArg-----AsnIleAsnLeuGlnSer	305
QY	736	CTGGGAATATTAAACGGCATCATCGACGAGCGATCCAGGCCCTTACTACCTGAATTC	795
Db	306	ValLeuIleGlyAsnGlyLeuThrAspGlyLeuThrGlnTyrGluTyrIleArgProMet	325
QY	796	GCTGTGAACAATACCTTACCGGTATPACAGGCTGTCAACGAGACCGCTACACACATCAAG	855
Db	326	AlaCysGlyAspGly---GlyTyrProAlaValLeuAspGlu---GlySerCysGlnAla	343
QY	856	TTTGCCAAACCAATGCTTCCAGGATTTGATTTCACCTGCAACAGACAAAC	915
Db	344	MetAspAsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerSerGlu	362
QY	916	CGCACCGCATTAGCTACTACGCGCTCTCGCGAAGCCACCAACATGTCCAGGACAAAT	975
Db	363	SerAla-----TrpValCysValProAlaSerIleTyrCysAsnAsnAla	377
QY	976	GTTGAGGGGCATATACGCTTGTGTTGCTGCGTGTGTATGATATTCGGCATCATAT	1035
Db	378	LeuLeuAlaProTyrGlnArg---ThrGlyGluAsnValTyrAspValArgGlyLysCys	396
QY	1036	GATGACCGC-----ACTCCGCCAAGTTATTACACAAATTTCTGGCAAG	1080
Db	397	GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys	416
QY	1081	GACTCTCTCATGAGCGCTATCGCGTCAACATCAAC---TACACCCAGTCCCAATATGAC	1137
Db	417	ThrGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp	436
QY	1138	GTCTACTACGCTTCTCCAGCAACAGGCGACTTTGCTGTGGCC-----AACTTCATCGAA	1191
Db	437	IleAsnArgAsnPheLeuPheHisGlyAspTyrMetLysProTyrHisArgLeuValPro	456
QY	1192	GACCTCGAGGAGATCTCTGCTCTCCCGCTCGCGTCTCTCTCATCTAT---GGCAGCGC	1248
Db	457	GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla	471
QY	1249	GATTACATCTGCAACTGTTCTGGCGGTGAGCGGCTTCTCCCTCGCTGCGAACTACTCCCA	1308
Db	472	AspPheIleCysAsnTyrLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTyrProGly	491
QY	1309	GCCGCCCATGTTCCGAAGCGGAGGTACACGCCCTCGAAAGTC-----AAC	1353
Db	492	GlnAlaGluTyrAlaSerAlaLysLeuGluAspLeuValValGluAsnGluHisLys	511
QY	1354	GGCTGCGAGTAGGGGAACCTCGGAGTAGTGGAATTTCTCTTCACTCCGCTCATGAG	1413
Db	512	GlyLysIysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly	531
QY	1414	CGAGCCATGAAGTCCCATACTACACGCCCATCCCTCCCTGCAATTTGTTTAAACCGACT	1473
Db	532	GlyGlyHisMetValProMetAspGlnProGluSerLeuGluPhePheAsnArgTrp	551
QY	1474	ATCTCGGT---TGG	1485
Db	552	LeuGlyGlyGluTrp	556

RESULT 11

US-08-608-224-4

: Sequence 4, Application US/08608224

: Patent No. 5205376



Db 229 PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla----- 244  
QY 436 AACCCITGAACGGGGTCGCGAAATTCAGAGCTTCAGGAGTTCAGGCGCGGTACCCA 495  
Db 244 ----- 244  
QY 496 ACCATTGATCCACTCTGATCATACCAATCTGCCGAGAGCGCGTGGGAGATC 555  
Db 245 -----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu 258  
QY 556 CTCGAGGATTCCTAGTGGACACACCTAGTGGAGCTCTAGGTCAGCTCAAGGACTTC 615  
Db 259 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaLysGlnAspPhe 274  
QY 616 AGTCTGAGCAGGAGCTAGGAGGCGACTATGCTCTCATCTCTCAATCATTTTAC 675  
Db 275 HisIleAlaGlySerTyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeu 294  
QY 676 GAGCAGAAAGAGAAATGCCAACGGTAGTGTAAATGCTTCAGCTTAATTTCAACTCT 735  
Db 295 SerHisLysLysArg-----AsnIleAsnLeuGlnSer 305  
QY 736 CTGGGAATTAATAACGGCATCATCGAGGCGGATCCAGGCGCCTTACTACCTGAATC 795  
Db 306 ValLeuIleGlyAsnGlyLeuThrAspGlyLeuThrGlnTyrGluTyrTyrArgProMet 325  
QY 796 GCTGTGCAACAATACCTACGCTATCAAGGCTGTCACGAGAGCGCTCACACTACATGAAG 855  
Db 326 AlaCysGlyAspGly---GlyTyrProAlaValLeuAspGlu---GlySerCysGlnAla 343  
QY 856 TTTCGCAACAATGCCAATGTTGCCAGGATTTGATTTCCACCTGCAACAGACAAAC 915  
Db 344 MetAspAsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerSerGlu 362  
QY 916 CGCAGCGCATTAGCTACCTACGCCCTCTCGCGGAGAGCCACCAACATGTCAGGGACAAT 975  
Db 363 SerAla-----TrpValCysValProAlaSerIleTyrCysAsnAsnAla 377  
QY 976 GTTAGGGGCCCATACTACGCCCTTGTGCTGCTGGGTGTATCATATTCGGATCCATAT 1035  
Db 378 LeuLeuAlaProTyrGluArg---ThrGlyGlnAsnValTyrAspValArgLysCys 396  
QY 1036 GATGACCCG-----ACTCGGCCAAGTATTACAAACAAATTTCTGGCAAG 1080  
Db 397 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 416  
QY 1081 GACTCTGTCATGAGCTATCGCGCTCAACATCAC---TACACCCAGTCCCAATAATGAC 1137  
Db 417 ThrGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp 436  
QY 1138 GTCTACTACGCTTTCAGCAACAGCGGACTTTGTGCGCCG-----AACTTCATCGAA 1191  
Db 437 IleAsnArgAsnPheLeuPheHisGlyAspTyrMetLysProTyrHisArgLeuValPro 456  
QY 1192 GACCTCGAGGAGATCCTGCTCTCCCGGTGGTCTCCCTCATCTAT---GGCGACGCC 1248  
Db 457 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla 471  
QY 1249 GATTACATCTGCACTGCTCGCGGCTGAGCGCGTTCCTCGCTGCGGCACTCCCAA 1308  
Db 472 AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTyrProGly 491  
QY 1309 GCGCCCGAGTTCGGAAGCGGGTACAGCGCCCTGAAAGTC-----AAC 1353  
Db 492 GlnAlaGluTyrAlaSerAlaLysLeuGluAspLeuValValGluAsnGluHisLys 511  
QY 1354 GCGTCGAGATGAGGGAACCTCGGAGTATGTTTCTCTTCTCCTCGGCTCTATGAG 1413  
Db 512 GlyLysLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly 531  
QY 1414 GCAGGCCATGAGTCCCATATACAGCCCATCGCTCTCCCTGCAATTTTAAACCGGACT 1473

Db 532 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPhePheAsnArgTrp 551  
QY 1474 ATCTTCGGT---IGG 1485  
Db 552 LeuGlyGlyGluTrp 556  
RESULT 12  
US-08-967-149-4  
; Sequence 4, Application US/08967149  
; Patent No. 5939305  
; GENERAL INFORMATION:  
; APPLICANT: Yaver, Debbie Sue  
; APPLICANT: Thompson, Sheryl Ann  
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF  
; TITLE OF INVENTION: ASPERGILLIUS NIGER  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: No. 5939305 No. 5939305disk of No. 5939305th America, Inc.  
; STREET: 405 Lexington Avenue, Suite 6400  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/967,149  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/608,452  
; FILING DATE: 28-FEB-1996  
; APPLICATION NUMBER: US 08/309,341  
; FILING DATE: 20-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lowney, Karen A.  
; REGISTRATION NUMBER: 31,274  
; REFERENCE/DOCKET NUMBER: 4247.000-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 867 0123  
; TELEFAX: 212 867 0298  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 557 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Aspergillus Niger  
; US-08-967-149-4  
Alignment Scores:  
Pred. No.: 2,9e-51 Length: 557  
Score: 623.00 Matches: 155  
Percent Similarity: 48.25% Conservative: 79  
Best Local Similarity: 31.96% Mismatches: 175  
Query Match: 21.45% Indels: 76  
DB: 2 Gaps: 18  
US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-08-967-149-4 (1-557)  
QY 79 GCAACAATGTCACCATCGGTACAGAACCCGGGCGAGAGCGCTTCGCGAGACTACC 138  
Db 132 AlaTyrAspLeuArgValLysThrAspProSerSerLeuGlyIle-----Asp 148  
QY 139 CCGGGTGCATACCTACTCTGATGATGTCACACTCTCCG---GAGTCCCATACCTTC 195  
Db 149 ProGlyValLysGlnTyrThrGlyTyrLeuAspAsnGluAsnAspLysHisLeuPhe 168



Query Match: 19.58% Indels: 85  
DB: 1 Gaps: 14

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-09-640-305-4 (1-491)

QY 79 GCAACAAATGTCACCATCCGCTACAGAAACCCGGGGCAGAGGGCGTTCGCGAGACTACC 138  
DB 62 AlaTyrSerLeuAlaGlyProLeuAspProLysSerLeuGlyVal----- 77

QY 139 CCGGTGTCAATCCTACTCTGGATATGTCGACACCTCTCCGAGTCCCATACCTTCCTC 198  
DB 78 AspThrValLysGlnIlePheSerGlyTyrLeuAspTyrGlnAspSerLysHisPhePheTyr 97

QY 199 TGGTCTTCGAGCAGACATACCCAGAACTCACATATGCTGCTTGAATGCT 258  
DB 98 TrpPheGluSerArgAsnAspProGluAsnAspProValIleLeuTyrLeuAsnGly 117

QY 259 GGCCTCGAAGCGATCTTTGATCGGTCTCTTCAAGAGTGGGCCCTTGCATGTCAAT 318  
DB 118 GlyProGlyCysSerSerPheValGlyLeuPheGluLeuGlyProSerSerIleGly 137

QY 319 TCGACTTTTGTGATGACTACATCAACCCCTCACTCGTGGAGAGGCTCCCAATTTACTATTC 378  
DB 138 AlaAspLeuLysProIleTyrAsnProTyrSerTrpAsnSerAsnAlaSerValIlePhe 157

QY 379 CTGTCCCGCAGCATGGGAGTCGGCTTTTCATATAGTATGAGGTGATGGTCCATTAAAC 438  
DB 158 LeuAspGlnProValGlyValGlyPheSerTyrGlyAsp----- 170

QY 439 CCTGTAAGTGGGTCGTCGAAATTCGAGCTTTCAGGAGTTCAGGCCCGGTACCCCAACC 498  
DB 170 ----- 170

QY 499 ATTGATGCCACTCTGATCGATACACTACCAATCTTCGCGCAGAGCGCGTTCGGAGATCTG 558  
DB 171 -----SerLysValSerThrThrAspAspAlaAlaLysAspValIlePheLeu 187

QY 559 CAAGGATTCCTAGTGGACTACTAGCTGGAGCTAGGGTGGAGCTTAAGGACTTCCAGT 618  
DB 188 AspLeuPhePheGluArgPheProHisLeu-----ArgAsnAspPheHis 203

QY 619 CTATGGCAGCAGGATGAGGCGCACTATGGTCCTGCAATCTTCAATCAATTTTACGAG 678  
DB 204 IleSerGlyGluSerTyrAlaGlyHisTyrLeuProLysIleAlaHis----- 219

QY 679 CAGAATGAGAAATGCCAACGGTAGTGTAAATGGTGTTCAGCTTAATTTCAACTCTCIG 738  
DB 220 -----GluIleAlaValValHisAlaGluAspSerSerPheAsnLeuSerSerVal 236

QY 739 GGAATTTAAGCGCATCATCGACAGGCGCATCCAGCGCCCTTACTPACCCCTGAATTCGCT 798  
DB 237 LeuIleGlyAsnGlyPheThrAspProLeuThrGlnTyrGlnTyrGluProMetAla 256

QY 799 GTGAACAATACCTACGGTAACAAGCTGTCACAGCAGCGCTACACTACATGAAGTTT 858  
DB 257 CysGlyGluGly--GlyTyrProAlaValLeuGlu----- 267

QY 859 GCCAACCAAAATGCCAAATGGTTGCCAGGAT----- 888  
DB 268 -----ProGluAspCysLeuAspMetAsnArgAsnLeuProLeuCysLeuSer 283

QY 889 TTGATTTCCACTGCAACACAGAACACCGCAGCATTAAGTACTAGCCCTCTGCGCC 948  
DB 284 LeuValAspArgCysTyrLysSerHis-----SerValPheSerCysVal 298

QY 949 GAAGCCACCAACATGTCAGGACACAAATGTGAGGGGCCATACACTACGCTTTTCTGGTGT 1008  
DB 299 LeuAlaAspArgTyrCysGluGlnGlnIleThrGly---ValTyrGluLysSerGlyArg 317

QY 1009 GGTGTGTATGATATTCGG-----CATCATATGATGACCCG-----ACT 1047  
DB 318 AsnProTyrAspIleArgSerLysCysGluAlaGluAspSerSerGlyAlaCysTyrGln 337

QY 1048 CCGCCAAAGTTATTACAAACAATTTCTGGCAAAGGACTCTGTCTATGACGACTATCGCGTTC 1107  
DB 338 GluGluIleTyrIleSerAspTyrLeuAsnGlnGluValGlnArgAlaLeuGlyThr 357

QY 1108 AACATCAACTACACCCAGTCC---AATAATGACGCTCTACTACGCTTTCAGCAACAGGC 1164  
DB 358 AspValSerSerPheGlnGlyCysSerSerAspValGlyIleGlyPheAlaPheThrGly 377

QY 1165 GACITGTCTGCCCAACTCTCAICGAGACCTCGAGGAGATCTTGTCTCCCTCCCGTGGT 1224  
DB 378 AspGlyProSerPro---PheHisGlnTyrValAlaGluLeuLeuAspGlnAspIleAsn 396

QY 1225 GTCCTCCCTCACTATGCGGAGCCGCTTACATCTCACTGCTGCGGGTTCAGCCGCTT 1284  
DB 397 ValLeuIleTyrAlaGlyAspLysAspTyrIleCysAsnTyrLeuGlyAsnLeuAlaTyr 416

QY 1285 TCCCTCGTGTGCGAATCTACTCCCAAGCCGCCAGTTCGCAAGCGGAGGTACAGCCCTG 1344  
DB 417 ThrGluLysLeuGluTyrArgTyrAsnGluGluTyrIleLysGlnValLeuArgThrTrp 436

QY 1345 AAAGTCAACGGCGTCGAG-----TATGGGAAACTCGCGAGTATGTAATTTCCCTTC 1398  
DB 437 LysSerGluGluThrAspGluThrIleGlyGluThrLysSerTyrGlyProLeuThrTyr 456

QY 1399 ACTCGCTCTATGAGCGAGCGCATGAGTCCCATACTACCCATCGCCCTCCCTGCAA 1458  
DB 457 LeuArgIleTyrAspAlaGlyHisMetValProHisAspGlnProGluAsnSerLeuGln 476

QY 1459 TTGTTTAAACGGGACTATC 1476  
DB 477 MetValAsnSerTrpIle 482

RESULT 14  
US-08-360-673-4  
; Sequence 4, Application US/08360673  
; Patent No. 5679544  
; GENERAL INFORMATION:  
; APPLICANT: Fleer, Reinhard  
; APPLICANT: Fournier, Alain  
; APPLICANT: Yeh, Patrice  
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR  
; PREPARATION AND USE  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Rd. 3043  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19002  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/360,673  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/FR93/00623  
; FILING DATE: 23-JUN-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 92/07785  
; FILING DATE: 25-JUN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, Julie K.  
; REGISTRATION NUMBER: 38,619  
; REFERENCE/DOCKET NUMBER: ST92040-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (610)454-3839  
; TELEFAX: (610)454-3808  
; INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 491 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-360-673-4

Alignment Scores:  
 Pred. No.: 491  
 Score: 568.50  
 Percent Similarity: 44.65%  
 Best Local Similarity: 31.07%  
 Mismatches: 184  
 Indels: 85  
 Gaps: 14  
 DB: 1

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-08-360-673-4 (1-491)

```

QY 79 GCAACAAATGTCACCATCGGTACAGGAACCGGGGAGAGGGCTCTCGAGACTACC 138
Db 62 AlaTyrSerLeuArgIleLysProLeuAspProLysSerLeuGlyVal----- 77
QY 139 CGGGTGTCAATCTACTCTGATGATGATGATGATGATGATGATGATGATGATGAT 198
Db 78 AspThrValLysGlnTrpSerGlyTyrLeuAspTyrGlnAspSerLysHisPheTyr 97
QY 199 TGGTCTTCGAGCCAGCAGACATAACCCAGAACTGACCTATCATCATTTGGTGAATGGT 258
Db 98 TrpPheGluSerArgAsnAspProGluAsnAspProValIleLeuTrpLeuAsnGly 117
QY 259 GGCCTCGGAGGATCTTTGATCGTCTCTCTCGAAGAGTGGGCGCTTCGATGTCAT 318
Db 118 GlyProGlyCysSerPheValGlyLeuPheGluLeuGlyProSerSerIleGly 137
QY 319 TCGATTTTTCATGACATACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 378
Db 138 AlaAspLeuLysProIleTyrAsnProTyrSerTrpAsnSerAsnAlaSerValIlePhe 157
QY 379 CTCTCCAGCATTGGAGTGGCTTTTTCATATAGTACATACGGTGTGATGGTCCATTAAC 438
Db 158 LeuAspGlnProValGlyValGlyPheSerTyrGlyAsp----- 170
QY 439 CCTGTAACTGGGTCTGTCGAAATTCGAGTTCGAGGAGTTCAGGGCGGTACCCCAACC 498
Db 170 ----- 170
QY 499 ATTGATCCCACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 558
Db 171 -----SerLysValSerThrThrAspAspAlaAlaLysAspValTyrIlePheLeu 187
QY 559 CAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGGTGCAGTCTAAGGACITCAGT 618
Db 188 AspLeuPheGluArgPheProHisLeu-----ArgAsnAspPheHis 203
QY 619 CTATGGAGGAGAGCTATGAGGAGCCTATGCTCTGATCTCTCAATTCATCACTTACGAG 678
Db 204 IleSerGlyGluSerTyrAlaGlyHisTyrLeuProLysIleAlaHis----- 219
QY 679 CAGATGAGAGCAATGTCACAGGTAGTGTAAATGTTGATGTTGATGTTGATGTTGAT 738
Db 220 -----GluIleAlaValValHisAlaGluAspSerSerPheAsnLeuSerVal 236
QY 739 GGAATTTATTAAGGCAATCATCGAGGCGGATCCAGGCCCTTACTACCTCAATTCGCT 798
Db 237 LeuIleGlyAsnGlyPheThrAspProLeuThrGlnTyrGlnTyrGluProMetAla 256
QY 799 GTGACAAATACCTACGGTATCAGGCTCTCAGGAGCCGCTCTACACTACATGAGTTT 858
Db 257 CysGlyGluGly---GlyTyrProAlaValLeuGlu----- 267
QY 859 GCCACCAATGCCAAATGGTTCGACGAT----- 888
Db 268 -----ProGluAspCysLeuAspMetAsnArgAsnLeuProLeuCysLeuSer 283

```

```

QY 889 TTGATTTCCACTGCAACAGACAAACCGCCAGCATTAAGTACTACGCCCTCTGCGCC 948
Db 284 LeuValAspArgCysTyrLysSerHis-----SerValPheSerCysVal 298
QY 949 GAAGCCACCAACATGTCAGGAGCAATGTTAGGGGCCCATACCTACGCTTCTGTCGT 1008
Db 299 LeuAlaAspArgTyrCysGluGlnIleThrGly---ValTyrGluLysSerGlyArg 317
QY 1009 GGTGTGTATGATATCGG-----CATCAATATGATGACCCG-----ACT 1047
Db 318 AsnProTyrAspIleArgSerLysCysGluAlaGluAspAspSerGlyAlaCysTyrGln 337
QY 1048 CGCCCAAGTTATTACACAAATTCGCAAGAGACTCTGTATGACGCTATCGCGGCTC 1107
Db 338 GluGluIleTyrIleSerAspTyrLeuAsnGlnGluValGlnArgAlaLeuGlyThr 357
QY 1108 ACATCAACTACACCCAGTCC---AATAATAGCTCTACTACGCTTTCCAGCAACAGCC 1164
Db 358 AspValSerSerPheGlnGlyCysSerSerAspValGlyIleGlyPheAlaPheThrGly 377
QY 1165 GACTTTGTCGCGCCCAACTTCATCGAAGACCTCGAGAGATCCTTGTCTCTCCCGTGGT 1224
Db 378 AspGlyProSerPro---PheHisGlnTyrValAlaGluLeuLeuAspGlnAspIleAsn 396
QY 1225 GTCTCCTCATCTATGCGGAGCGCGATTACATCTGCAACTGGTTCGGCGGTCTAGCGGCT 1284
Db 397 ValIleTyrAlaGlyAspLysAspTyrIleCysAsnTrpLeuGlyAsnLeuAlaTrp 416
QY 1285 TCCTCTCGTCCGAATACTCCCAAGCGCGCGGTTCGGAAGCGCGGTACACGCCCTG 1344
Db 417 ThrGluLysLeuGlnTrpArgTyrAsnGluTyrLysLysGlnValLeuArgThrTrp 436
QY 1345 AAGTCAACCGGCTCGAG-----TATGGGAACCTCGCGATGATGATGATGATGATG 1398
Db 437 LysSerGluGluThrAspGluThrIleGlyLeuThrLysSerTyrGlyProLeuThrTrp 456
QY 1399 ACTCGGCTCTATGAGCGAGCGCATGAAGTCCCATCTACACGCCCATCGCTCCCTGCA 1458
Db 457 LeuArgIleTyrAspAlaGlyHisMetValProHisAspGlnProGluAsnSerLeuGln 476
QY 1459 TTGTTTAAACCGGACTATC 1476
Db 477 MetValAsnSerTrpIle 482

```

RESUL 15  
 US-08-899-324-33  
 ; Sequence 33, Application US/08899324  
 ; Patent No. 5945329  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bredam, Klaus  
 ; APPLICANT: Keilland-Brandt, Morten  
 ; APPLICANT: Mortensen, Uffe  
 ; APPLICANT: Olesen, Kjeld  
 ; APPLICANT: Stennicke, Henning  
 ; APPLICANT: Wagner, Fred  
 ; TITLE OF INVENTION: CUSTOMIZED PROTEASES  
 ; NUMBER OF SEQUENCES: 33  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
 ; STREET: 3100 No. 5945329west Center, 90 S. 7th Street  
 ; CITY: Minneapolis  
 ; STATE: MN  
 ; COUNTRY: U.S.A.  
 ; ZIP: 55402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq Version 1.5  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/899,324  
 ; FILING DATE: 23-JUL-1997  
 ; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/329,892  
 ; FILING DATE: 27-OCT-1994  
 ; APPLICATION NUMBER: 08/144,704  
 ; FILING DATE: 28-OCT-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kettleberger, Denise M  
 ; REGISTRATION NUMBER: 33,924  
 ; REFERENCE/DOCKET NUMBER: 8648.44USC1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 612/332-5300  
 ; TELEFAX: 612/332-9081  
 ; TELEX:

; INFORMATION FOR SEQ ID NO: 33:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 532 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; FRAGMENT TYPE: Internal  
 ; ORIGINAL SOURCE:  
 ; US-08-899-324-33

Alignment Scores:  
 Pred. No.: 1,1e-41 Length: 532  
 Score: 523.50 Matches: 147  
 Percent Similarity: 43.68% Conservative: 74  
 Best Local Similarity: 29.05% Mismatches: 200  
 Query Match: 18.03% Indels: 85  
 DB: 2 Gaps: 19

US-09-712-338-1\_COPY\_55\_1662 (1-1608) x US-08-899-324-33 (1-532)

QY 43 CCGAAGACCCACCGGGGTCAAGACT-----CITACAAACCGCAACAACT 87  
 Db ||||| ||| ::||| |||||  
 QY 82 ProLysPheProGluAlaIleLysThrLysLysAspTrpAspPheValValLysAsnAsp 101  
 Db ||||| |||||  
 QY 88 GTCACCATCCGGTAC-----AAGGAACCGGGGCGAGAGGGCGTC 126  
 Db ||||| |||||  
 QY 102 AlaIleGluAsnTyrGlnLeuArgValAsnLysIleLysAspProLysIleLeuGlyIle 121  
 Db ||||| |||||  
 QY 127 TCGAGACTACCCGGGTGTCAAATCTCTGATATGTCGACACCTCCCGAG--- 183  
 Db ||||| |||||  
 QY 122 -----AspProAsnValThrGlnTyrThrGlyTyrLeuAspValGluAspGluAsp 138  
 Db ||||| |||||  
 QY 184 TCCCATACCTCTCTCTGTTCTCGAAGCCAGACATAACCCAGAACTGCACCTATCACA 243  
 Db ||||| |||||  
 QY 139 LysHisPhePheThrPheGluSerArgAsnAspProAlaLysAspProValIle 158  
 Db ||||| |||||  
 QY 244 TTGGTGTGAATGTCGCCCTGGAAGCGATTTCTTGATCGTCTCTTGAAGAGTGGGC 303  
 Db ||||| |||||  
 QY 159 LeuTrpLeuAsnGlyGlyProGlyCysSerSerLeuThrGlyLeuPheGluLeuGly 178  
 Db ||||| |||||  
 QY 304 CTTGCCATGTCAATTCGACTTTTGATCACTACATCAACCTCACTCTGTCGAAACGAGTC 363  
 Db ||||| |||||  
 QY 179 ProSerSerIleGlyProAspLeuLysProIleGlyAsnProTyrSerTrpPaspSerAsn 198  
 Db ||||| |||||  
 QY 364 TCCAAATTACTATCTCTGCCACCATTTGGAGTCGGCTTTTCATATAGTATGATACGGTT 423  
 Db ||||| |||||  
 QY 199 AlaThrValIlePheLeuAspGlnProValAsnValGlyPheSerTyrSer----- 215  
 Db ||||| |||||  
 QY 424 GATGGGTCCATTAAACCTGTAACTGGGGTCTCGCAAAATCGAGCTTTGCGAGGAGTTCAG 483  
 Db ||||| |||||  
 QY 216 -----GlySerSer 218  
 QY 484 GCGCGGTACCCACCAATTGATGCCACTCTGATCGATACCTACCAATCTGCCGCGAGAGCC 543  
 Db ||||| |||||  
 QY 219 Gly-----ValSerAsnThrValAlaAlaGlyLysAsp 229  
 QY 544 GCTTGGGAGATCTCGAAGGATTCCTTAGTGGACTACCTAGCTTGGACICTAGGGTGCAG 603

Db 230 ValTyrAsnPheLeuGluLeuPhePheAspGlnPheProGluTyrValAsnLysGlyGln 249  
 QY 604 TCTAAGCACTTCAGTCTATGGAGCGAGAGCTATGGAGGCGACACTATGCTCTGCATTTCTTC 663  
 Db ||||| |||||  
 QY 250 -----AspPheHisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheAla 267  
 QY 664 AATCAATTTTACGAGCAGAAATGAGAAATGCGCAACGGTAGTGTAAIGTGTTCAGCTT 723  
 Db ||||| |||||  
 QY 268 SerGluIleLeuSerHisLysAspArg----- 276  
 QY 724 AATTCAACTCTCTGGGAATTATT-----AACGGCATCATCGACAGGCGATCCAGGCC 777  
 Db ||||| |||||  
 QY 277 AsnPheAsnLeuThrSerValIleLeuGlyAsnGlyLeuThrAspProLeuThrGlnTyr 296  
 QY 778 CCTTACTACCTGAATTGCGTGTGAACAATACCTAGGTATCAAGCTGTCACAGAGACC 837  
 Db ||||| |||||  
 QY 297 AsnTyrTyrGluProMetAlaCysGlyGluGly--GlyGluProSerVal----- 312  
 QY 838 GCTTCACTACATGATGAAGTTTGCACCAATGCCAAATGGTTGCCAGGATTTGATTTCC 897  
 Db ||||| |||||  
 QY 313 -----LeuProSerGluGluCysSerAlaMetGluAspSerLeuGlu 326  
 QY 898 ACCTGCAACAGACACAAACCGCACCTTACCTAGCTACCTACGCCCTC-----TGCGCCGA 951  
 Db ||||| |||||  
 QY 327 ArgCysLeuGlyLeuIleGluSerSerTyrAspSerGlnSerValTrpSerCysValPro 346  
 QY 952 GCCACCAACATGTGCGAGGCAATGTGAGGGGCCATACTAGCCCTTCTGCTGCTGGTGGT 1011  
 Db ||||| |||||  
 QY 347 AlaThrIleTyrCysAsnAlaGlnLeuAlaProTyrGlnArg---ThrGlyArgAsn 365  
 QY 1012 GTGTATGATTTCCGGCATCCATATGATGAC-----CCGACTCCGCCAAGT 1056  
 Db ||||| |||||  
 QY 366 ValTyrAspIleArgLysAspCysGluGlyGlyAsnLeuCysTyrProThrLeuGlnAsp 385  
 QY 1057 TATTACACAAATTTCTGCCAAGAGCTCTGTATGAGCGCTATCGCGCTCAACATCAAC 1116  
 Db ||||| |||||  
 QY 386 ---IleAspAspTyrIleuAsnGlnAspTyrValLysGluAlaValGlyAlaGluValAsp 404  
 QY 1117 TACACCCAGTCC---AATAATCAGCTCTACTACGCTTTCACCAACAGCGACITTTGC 1173  
 Db ||||| |||||  
 QY 405 HisTyrGluSerCysAsnPheAspIleAsnArgAsnPheLeuPheAlaGlyAspTrpMet 424  
 QY 1174 TGGCCCACTTCTCATGAAAGACCTCGAGGAGATCTTGTCTCTCCGCTGCGTGTCTCCCTC 1233  
 Db ||||| |||||  
 QY 425 LysPro---TyrHisThrAlaValThrAspLeuLeuAsnGlnAspLeuProIleLeuVal 443  
 QY 1234 ATCTATGGCGAGCGCGATACATCTGCAACCTGTTGCGGGGTCAGCGCTTCCCTCGCT 1293  
 Db ||||| |||||  
 QY 444 TyrAlaGlyAspLysAspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrAspVal 463  
 QY 1294 GCGAACTACTCCCAAGCGCGCTTCCGAAGC-----GCAGGTACAGCCCGCTG 1344  
 Db ||||| |||||  
 QY 464 LeuProTrpLysTyrAspGluPheAlaSerGlnLysValArgAsnTrpThrAlaSer 483  
 QY 1345 AAAGTCACGGCGTCGAGTATGCGGAAACTCCGAGTATGGTAATTTCTCTTCTACCTGCG 1404  
 Db ||||| |||||  
 QY 484 IleThrAspGluValAla---GlyGluValLysSerTyrLysHisPheThrTyrLeuArg 502  
 QY 1405 GTCTATGAGCGAGCGCATGAAGTCCCATCTACAGCCCATCGCTCCCTCGAATTTGTTT 1464  
 Db ||||| |||||  
 QY 503 ValPheAsnGlyGlyHisMetValProPheAspValProGluAsnAlaLeuSerMetVal 522  
 QY 1465 AACCGGACTATCTTCGGT 1482  
 Db ||||| |||||  
 QY 523 AsnGluTrpIleHisGly 528

Search completed: September 16, 2003, 18:49:45  
 Job time : 51 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 22:50:17 ; Search time 54 Seconds  
(without alignments)  
1480.662 Million cell updates/sec

Title: US-09-712-338-2\_COPY\_19\_555

Perfect score: 2887

Sequence: 1 LFGSTPASVGRRLPKNPTG.....HTQSSVPLPLTATSMSSVGMA 537

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 556269 seqs, 148893369 residues

Total number of hits satisfying chosen parameters: 556269

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	517.5	17.9	421	9	US-09-420-785A-4
2	517.5	17.9	421	9	US-09-901-252-15
3	361	12.5	476	14	US-10-084-018-3
4	359	12.4	476	10	US-09-909-320-164
5	359	12.4	476	10	US-09-909-088B-164
6	359	12.4	476	10	US-09-905-291A-164
7	359	12.4	476	10	US-09-902-853-164
8	359	12.4	476	10	US-09-907-824-164
9	359	12.4	476	10	US-09-907-841-164
10	359	12.4	476	11	US-09-904-011-164
11	359	12.4	476	11	US-09-906-742-164
12	359	12.4	476	11	US-09-906-838-164
13	359	12.4	476	11	US-09-907-613-164
14	359	12.4	476	11	US-09-907-942-164
15	359	12.4	476	11	US-09-796-753-40

16	359	12.4	476	11	US-09-904-859-164	Sequence 164, App
17	359	12.4	476	11	US-09-909-204-164	Sequence 164, App
18	359	12.4	476	11	US-09-904-820-164	Sequence 164, App
19	359	12.4	476	11	US-09-904-786-164	Sequence 164, App
20	359	12.4	476	11	US-09-906-646-164	Sequence 164, App
21	359	12.4	476	11	US-09-906-700-164	Sequence 164, App
22	359	12.4	476	11	US-09-903-786-164	Sequence 164, App
23	359	12.4	476	11	US-09-902-903-164	Sequence 164, App
24	359	12.4	476	11	US-09-903-749A-164	Sequence 164, App
25	359	12.4	476	11	US-09-904-119-164	Sequence 164, App
26	359	12.4	476	11	US-09-904-956-164	Sequence 164, App
27	359	12.4	476	11	US-09-902-736-164	Sequence 164, App
28	359	12.4	476	11	US-09-907-794-164	Sequence 164, App
29	359	12.4	476	11	US-09-903-943-164	Sequence 164, App
30	359	12.4	476	11	US-09-904-463-164	Sequence 164, App
31	359	12.4	476	11	US-09-907-925-164	Sequence 164, App
32	359	12.4	476	11	US-09-902-692-164	Sequence 164, App
33	359	12.4	476	11	US-09-903-520-164	Sequence 164, App
34	359	12.4	476	11	US-09-905-056-164	Sequence 164, App
35	359	12.4	476	11	US-09-909-064-164	Sequence 164, App
36	359	12.4	476	11	US-09-904-553-164	Sequence 164, App
37	359	12.4	476	11	US-09-905-381-164	Sequence 164, App
38	359	12.4	476	11	US-09-905-088-164	Sequence 164, App
39	359	12.4	476	11	US-09-907-575-164	Sequence 164, App
40	359	12.4	476	11	US-09-905-075-164	Sequence 164, App
41	359	12.4	476	11	US-09-902-759-164	Sequence 164, App
42	359	12.4	476	11	US-09-902-634-164	Sequence 164, App
43	359	12.4	476	11	US-09-902-713-164	Sequence 164, App
44	359	12.4	476	11	US-09-907-979-164	Sequence 164, App
45	359	12.4	476	11	US-09-902-615-164	Sequence 164, App

#### ALIGNMENTS

RESULT 1  
US-09-420-785A-4  
; Sequence 4, Application US/09420785A  
; Patent No. US20010010923A1  
; GENERAL INFORMATION:  
; APPLICANT: MORTENSEN, UFFE  
; APPLICANT: OLESEN, KJELD  
; APPLICANT: STENNICKE, HENNING  
; APPLICANT: SØRENSEN, SIEEN B.  
; APPLICANT: BREDDAM, KLAUS  
; TITLE OF INVENTION: MODIFIED CARBOXYPEPTIDASE  
; FILE REFERENCE: 089187/0109  
; CURRENT APPLICATION NUMBER: US/09/420,785A  
; CURRENT FILING DATE: 1999-10-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 421  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-420-785A-4

Query Match 17.9%; Score 517.5; DB 9; Length 421;  
Best Local Similarity 29.5%; Pred. No. 1.5e-40;  
Matches 142; Conservative 70; Mismatches 185; Indels 85; Gaps 17;  
QY 33 RYKPEAGVCGTTPGVKXSGYVDTSPF-SHTFFWFARHNPETAPITLWNGGSGD 91  
Db 1 KIKDKILGI--DPNVTQYTGILDVEDKHFHFWFESRNDPAKDPVILWNGGPGCS 57  
QY 92 SLTGLFELGFCVHNSTFDYINPHSWNEVSNLFLSQPLGVGFYSYDVTGDSINPTGV 151  
Db 58 SLTGLFELGFCVHNSTFDYINPHSWNEVSNLFLSQPLGVGFYSYDVTGDSINPTGV 104  
QY 152 VENSFPAGVGPRYPTIDATLIDTNTLAAEAWEILQGLSLGSLDSRVSKDFSLWTES 211  
Db 105 -----GSSG-----VSNTVAGKDVYNFLELFFDQFPFVYVNGQ--DFHTAGES 146

QY 212 YGCHYPAFNFHFEQNERIANGSVNGVQLNFNSLGI--NGIIDEAIOAPYYPEFA-- 266  
 DB 147 YAGHYLPVFASEILSHKDR-----NFNLTSVLIGNGLTDPLOINYYEPMACGE 195  
 QY 267 -----VNNTYGIKAVNEVTYNYMKFANOMPNGCODLSTCKQNRNALADYALCAEATNM 321  
 DB 196 GGEPSVLPSECSAMEDSL-----ERCLGLTESCYDSQ-----SWMSCVPATII 239  
 QY 322 CRNVEGPIYAFAGRGVYDIRHPYDD-----PTPPSYNKFELAKDSVMDAIGVNNVITQS 376  
 DB 240 CNAQAUPYQOR-TGRNVDYIRKCEGNNLCYPTIQD-IDDYLNQDYVKEAVGAERVDHYES 297  
 QY 377 -NNDVYAFQGTGFVMPNFTEDEEILALPVRVSLYGDADYICNMFEGGOAVSLAANY 435  
 DB 298 CNFIDINNELFAGDWMKP-YHTAVTDLLNQDLPILVYAGDKDFICNWLGNKAWTDVLPWK 356  
 QY 436 QAAQFRS---AGYTPLKVNVEYGETREYGNFSTRVYEAGHEVPYQPIASLQLFNRTI 492  
 DB 357 YDEEFASQKVRNWTASITDEVA-GEVKSYPKFTYLRVFNHGHHVDFDPENALSMVNEWI 415  
 QY 493 FG 494  
 DB 416 HG 417

RESULT 2

US-09-901-252-15  
 ; Sequence 15, Application US/09901252  
 ; Patent No. US2002026658A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chapple, Clint  
 ; TITLE OF INVENTION: Genes Encoding Sinapoylglucose:Malate Sinapoyltransferase and Met  
 ; FILE REFERENCE: N1422-004  
 ; CURRENT APPLICATION NUMBER: US/09/901,252  
 ; PRIOR FILING DATE: 2001-07-09  
 ; PRIOR APPLICATION NUMBER: US 60/216593  
 ; PRIOR FILING DATE: 2000-07-07  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: Patent version 3.0  
 ; SEQ ID NO 15  
 ; LENGTH: 421  
 ; TYPE: PRT  
 ; ORGANISM: Saccharomyces  
 US-09-901-252-15

Query Match 17.9%; Score 517.5; DB 9; Length 421;  
 Best Local Similarity 29.5%; Pred. No. 1.5e-40;  
 Matches 142; Conservative 70; Mismatches 185; Indels 85; Gaps 17;

QY 33 RYKEPGAEGVCTTPGVKSYSGYVDTSP-SHTFFWFFEARHPNPETAPITLWNGGPGSD 91  
 DB 1 KIKDPKILGI---DPNVTQTYGLDVEDEKHHFFWTFESRNDPAKDPVILWNGGPGS 57  
 QY 92 SLITGLPEELGPCHVNSTFDYINPHSWNEVSNLELSQLPGVGFYSVDIVGGINPVTGV 151  
 DB 58 SLTGLFELGPPSIGPDLKPIGNPYSWNSNATVIFLDQPVNVGFSYS----- 104  
 QY 152 VENSSTAGVQGRYPTIDAILDTITNLAAEAWEILQGLSLPSLRSVQSKDFSLWTES 211  
 DB 105 -----GSSG-----VSNTVAAGKDVYNFELFDQFPPEYVKNKG--DFHAGES 146  
 QY 212 YGCHYPAFNFHFEQNERIANGSVNGVQLNFNSLGI--NGIIDEAIOAPYYPEFA-- 266  
 DB 147 YAGHYLPVFASEILSHKDR-----NFNLTSVLIGNGLTDPLOINYYEPMACGE 195  
 QY 267 -----VNNTYGIKAVNEVTYNYMKFANOMPNGCODLSTCKQNRNALADYALCAEATNM 321  
 DB 196 GGEPSVLPSECSAMEDSL-----ERCLGLTESCYDSQ-----SWMSCVPATII 239  
 QY 322 CRNVEGPIYAFAGRGVYDIRHPYDD-----PTPPSYNKFELAKDSVMDAIGVNNVITQS 376  
 DB 240 CNAQAUPYQOR-TGRNVDYIRKCEGNNLCYPTIQD-IDDYLNQDYVKEAVGAERVDHYES 297

QY 377 -NNDVYAFQGTGFVMPNFTEDEEILALPVRVSLYGDADYICNMFEGGOAVSLAANY 435  
 DB 298 CNFIDINNELFAGDWMKP-YHTAVTDLLNQDLPILVYAGDKDFICNWLGNKAWTDVLPWK 356  
 QY 436 QAAQFRS---AGYTPLKVNVEYGETREYGNFSTRVYEAGHEVPYQPIASLQLFNRTI 492  
 DB 357 YDEEFASQKVRNWTASITDEVA-GEVKSYPKFTYLRVFNHGHHVDFDPENALSMVNEWI 415  
 QY 493 FG 494  
 DB 416 HG 417

RESULT 3

US-10-084-018-3  
 ; Sequence 3, Application US/10084018  
 ; Publication No. US20020160499A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bandman, Olga  
 ; Hawkins, Phillip R.  
 ; Hillman, Jennifer L.  
 ; Lal, Preeti  
 ; Goli, Surya K.  
 ; TITLE OF INVENTION: NOVEL HUMAN SERINE  
 ; CARBOXYPEPTIDASE  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/084,018  
 ; FILING DATE: 25-Feb-2002  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/299,689A  
 ; FILING DATE: <Unknown>  
 ; APPLICATION NUMBER: 08/828,488  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J.  
 ; REGISTRATION NUMBER: 36,749  
 ; REFERENCE/DOCKET NUMBER: PF-0241 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-855-0555  
 ; TELEFAX: 415-845-4166  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 476 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: MMLR3DT01  
 ; CLONE: 566993  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
 US-10-084-018-3

Query Match 12.5%; Score 361; DB 14; Length 476;  
 Best Local Similarity 26.7%; Pred. No. 1.5e-25;  
 Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;  
 QY 47 PG--VKYSYG--VDTSPSTHTFFWFFEARHPNPETAPITLWNGGPGSDSLIGLFEELGP 102  
 DB 67 PGLNMKSYAGPLTVNKTNSNLTFFWFFPAQIQEDAPVYLWLOGGPGSSMXGLFVEHGP 126



Qy	103	CHVNSTFD	DDY	INPHSWNEVSNLLFLSQPLGVGFYSYSDTV	DG	SINPVTGVVENS	FAGVQ	162			
Db	127	YVVT	SNMFLRDRDP	FWTTXSM	LYIDNPVG	GFSTDDTHG	-----	167			
Qy	163	RYPIDA	LTIDTNLA	EAAEILLOFT	SLGPLSRV	QSKDPSLWTSY	GHHYGP	222			
Db	168	-----	YAVEDD	VAROLYSAL	IOFF-----	QIFPEYK	KNDFYTG	SYACKYVFAIAH 215			
Qy	223	HFYEQNER	ANGSVNGV	OLNFS	LIGIIN	IGIDEA	IQAPYEP	FAVNNYGIKAVNETVYN 282			
Db	216	LIHSLN	-----	PVREVK	INLNGA	IGDYS	DPES	IGGYAEF-----YQIGLLDEKOKK 265			
Qy	283	YMKFANQ	PNPCOD	LLSTCK	QOIN-----	RVALADYAL	CABAIN	-----MCRDNVGE--PYATAFAG 335			
Db	266	Y--FQKO	-----	CHCEI	EHIRKON	WEAF	FEILDK	LIDGLTSDPSYFQNVGCSNYNFI-- 318			
Qy	336	RGVYDIRH	YDDPP	PPSYN	KFLAKDS	YMDAIG	NYNIN	YQSNNDY--YAFQOTGD	FVWP 393		
Db	319	-----	RCTEP	EDQLYYK	FLS	PEVRQ	AIHVG--	NOTFNDG	ITIVKLYREDT	VQSKVP 369	
Qy	394	NFIEDLE	EIALPVR	SLIT	GDADYI	-----	CNWF	GGOAVSLA	ANYSAQA-- 439		
Db	370	WLTEIM	NN-----	YKVL	IYNGQLD	IIVAAAL	TERS	SLMGMD	WKGSQ-----EYKAEKV 418		
Qy	440	---FRS---	AGYPL	KVNGV	EYGETREY	GNFST	RV	VEACHE	VPYPIAS	LOLFNRI 492	
Db	419	WKLEKSD	SEVAGY-----	-----	IRQV	GD	HQVIL	IRGG	GHILFYD	QPLAF	DMINRFI 466
Qy	493	F--GWD	496								
Db	467	YKG	GWD 472								

## RESULT 4

US-09-909-320-164  
Sequence 164, Application 05/05909320  
Patent No. US20020132240A1  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, A.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kljavin, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: 10466-14  
CURRENT APPLICATION NUMBER: US/09/909,320  
CURRENT FILING DATE: 2002-01-04  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07

```

; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-909-320-164

Query Match          12.4%; Score 359; DB 10; Length 476;
Best Local Similarity 26.7%; Pred. No. 2.3e-25;
Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;

QY 47 PG--VKSYSGY--VDISPESHTEFFWFTEARHNPTATITLWNGSPGSDSLIGLFEEELGP 102
   |||::||: |::|||::||: |||::||: |||::|||::|||::|||::|||::|||::|||
Db 67 PGLNMKSYAGFLT VKNKYNSNLFWFPEAQLOPEDAPVWLQGGPGGSMFGLFVEHGP 126

QY 103 CHVNSTEDDYNPHSWNEVSNLLFLSOPLGVCFSYSTVDGSINPVTGVVENSSEFAGVQG 162
Db 127 YVVISNMTLRDRDPFWTITLSMLYIDNPVGTFGSFTDDTHG-----167

QY 163 RYPTIDATLDITNLAAEAAWEILLQGLSLPSLDSRVQSKDFSLWTESYGGHYGPFPN 222
   :::::||||: |||::||: |||::||: |||::||: |||::||: |||::||: |||::||:
Db 168 -----YANNEDDVARDLYSALIQQF-----QIFPEYKNNDFFVTTGESYAGYVPVAIAH 215

QY 223 HFYEQNRIANGSVNGVOLNFNSLIIGTINGIIDEAIOAPYPPEFAVNNTYGIKAVNETVYN 282
Db 216 LIHSLN-----PVREVKNLNGIAICDGYSDEPSIEIIGGYAEF----LYIGILLDEKQKK 265

QY 283 YMKFANOMPNGCQDLISTCKQTN---RTALADVALCAEATN--MCRDNVEG--PYAFAG 335
Db 266 Y--FQXQ-----CHECHETIRKWNFEAFELDKLLDGLTSDPSYEQNVGTCSNYNFL- 318

QY 336 RGVIDRHPYDDTPPSYNNKFIAKDSVMDAIGVINYNIGSNNDVY--YAFQQTGDVWP 393
   :|||::||: |||::||: |||::||: |||::||: |||::||: |||::||: |||::||:
Db 319 -----RCIEPEDQLYYVVFSLDEVKQAIIHG-NQTFNDGTIVEKYLREDTVQSVPK 369

QY 394 NTFEDLEELIALPVRSYLIYGDAFYI-----CNWFGQAVSLAANTSOAAQ-- 439
   :|||::||: |||::||: |||::||: |||::||: |||::||: |||::||: |||::||:
Db 370 WLTEIMNN-----YKVLIYNGQLDIIVAAALTERSLMGMWDKGSQ-----EFKKKEKKV 418

QY 440 ---FRS-----AGYTPLKVNCEYGETREYGNSTFTRYEAGHEVPYIPTYASIQLENRTI 492
   :|||::||: |||::||: |||::||: |||::||: |||::||: |||::||: |||::||:
Db 419 WKLFKSDSEVAGY-----IROAGDFHQVIIRGGGHILIPYDOLPLRAFDMINRFT 466

```



```

Db      266 Y--FQKQ---CHECIEHIRKQNWPEAFELDKLLDGLTSDPSYFQNYGCSYYNFI-- 318
QY      336 RGVYDIRHPYDPTPPSYNKFSLAKDSYMDALGVNINVTQSNDVY--YAFQQTGDFVWP 393
Db      319 -----RCTEPEDQLYVVKFLSLPEVROAIHVG--NOIFNDGTIVEKYLREDTVGSVKP 369
QY      394 NFIEDLEILPALPVRSILYGDADYI-----CNWFGGGAVALAANYSSAAQ-- 439
Db      370 WLTEIMNN-----YKVLINQGLDIIVAALTEKSLMGMDWKGSQ-----EYKRAEKV 418
QY      440 ---FRS-----AGYTPLKVNQVEYGETREYGNFSTFVVYEAAGHEVPYQPIASLQLFNRIT 492
Db      419 WKIFKSDSEVAGY-----IRQAGDFHQQVIIRGGGHILPYDQPLRAFDMINRFI 466
QY      493 F--GWD 496
Db      467 YGKGWD 472

RESULT 7
US-09-902-853-164
: Sequence 164, Application US/09902853
: Publication No. US20020192659A1
: GENERAL INFORMATION:
: APPLICANT: Genentech, Inc.
: APPLICANT: Ashkenazi, Avi
: APPLICANT: Botstein, David
: APPLICANT: Desnoyers, Luc
: APPLICANT: Patton, Dan L.
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Fong, Sherman
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Gerber, Hanspeter
: APPLICANT: Gerritsen, Mary E.
: APPLICANT: Goddard, A.
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, Christopher J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Hillan, Kenneth, J.
: APPLICANT: Kljavin, Ivar J.
: APPLICANT: Mather, Jennie P.
: APPLICANT: Pan, James
: APPLICANT: Paoni, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William, I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: FILE OF INVENTION: Acids Encoding the Same
: FILE REFERENCE: 10466-14
: CURRENT APPLICATION NUMBER: US/09/902,853
: PRIOR FILING DATE: 2001-07-10
: PRIOR APPLICATION NUMBER: US/09/665,350
: PRIOR FILING DATE: 2000-09-18
: PRIOR APPLICATION NUMBER: US 60/143,048
: PRIOR FILING DATE: 1999-07-07
: PRIOR APPLICATION NUMBER: US 60/145,698
: PRIOR FILING DATE: 1999-07-26
: PRIOR APPLICATION NUMBER: US 60/146,222
: PRIOR FILING DATE: 1999-07-28
: PRIOR APPLICATION NUMBER: PCT/US99/20594
: PRIOR FILING DATE: 1999-09-08
: PRIOR APPLICATION NUMBER: PCT/US99/20944
: PRIOR FILING DATE: 1999-09-13
: PRIOR APPLICATION NUMBER: PCT/US99/21090
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/21547
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/23089
: PRIOR FILING DATE: 1999-09-15

```





```

; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hauspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavich, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-904-011-164

Query Match 12.4%; Score 359; DB 11; Length 476;
Best Local Similarity 26.7%; Pred No. 2.3e-25;
Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;

QY 47 PG--VKSYGY--VDTSPESHTEFFFEARHNHETAPITLWNGPGSDSLIGFEELGP 102
Db 67 PGLMKSYAGFLTVNKTYSNLFNFFPAQIQEDAPVVLWLOGGPGSSMFLGFEHGP 126
QY 103 CHVNSTFDDYINPHSNWVSNLLFLSOPLGVSYSYDVGSDINPVIGVYVENSFAGVQ 162
Db 127 YVVTNNMLRDRDPWTTLISMLYIDNPVGTGFSFTDDTHG----- 167
QY 163 RYPTIDATLDTNLAARAAWEILQGLSLPILSDRSVQSKDSLTWSTESYGGHYGPAFFN 222
Db 168 -----YAVNEDDVARDLYSALIQFF-----QIFPEYKNDNFVVTGESAGKYVPAIAH 215
QY 223 HEYQNERIANGSVNGVQLNFNSLGIINGIIDEAIAQAPYYPEFAVNNTYGIKAVNETVYN 282
Db 216 LIHSLN-----PVREVKININGIAIGDYSDPESIIGGYAEF-----LYIGLIDDERQKK 265
QY 283 YMKFANOMPGQDLISTCKOTN---RTALADYALCAETN---MCRDNVEG--PYXAFAG 335
Db 266 Y--PQKQ---CHECIEHIRKQNWFEAFELDKLLDGLTSDPSYFQNVGTGCSNYNLF- 318
QY 336 RGVDIRHPYDDPTPPSYNKKFLAKDSVMDAIGVNIYNTQSNNDVY--YAFQQTCDFWVP 393
Db 319 -----RTEPEDQLYYVKFELSPEVRQALHVG--NOTENDGIVKEYLREDIVQSVKP 369
QY 394 NFIEDLEILLALPVRSVLIYGDADYI-----CNWFGQAVSLAANYSOAAQ-- 439
Db 370 WLTEIMNN-----YKVLINQGLDIIVAAALTSRLMGMWKGSQ-----EYKKAERKV 418
QY 440 ---FRS-----AGYTPLVNGVVEYGETREYGNFSTRVVEAGHEVPYQPIASLQLFNRTI 492
Db 419 WKIEKSDSEVAGY-----IROAGDFHQVIRGGHILLPYDQPLRAFADMINRFI 466
QY 493 F--QWD 496
Db 467 YKGWMD 472

RESULT 11
US-09-906-742-164
; Sequence 164, Application US/09906742
; Publication No. US20030023054A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hauspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavich, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
```

[illegible]

```

; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-906-838-164

Query Match      12.4%; Score 359; DB 11; Length 476;
Best Local Similarity 26.7%; Pred. No. 2.3e-25;
Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;

47 PG--VKSYSGY--VDTSPESTHTFFWFEARHNPETAPITLWLNKGPGSDSLICLFEELGP 102
   || :|||:| : : : :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
67 PGLNMKSYAGFLTVNKTYNSNLFWFPPAQIQPEDAPVVLWLOGGPGGSSMFLGFVEHGP 126

103 CHVNSTFDDYINPHSNWNEVSNLLFSLQPLGVGFSDIVDGSINPVTGVVENSFAQVQG 162
127 YVVTSMNLTDRDRFPWTTTSLMLYIDNPVGTGSFTDDTHG----- 167
163 RYPTIDATLDTINLAAEAAWEILOGLFSLGSLDSRVQSKDFSLWFSYGGHYGAFFN 222
168 -----YAVNEDDVARDLYSALIQFE-----QIFPEYKKNDDYVVTGESYAGYVPAIAH 215
223 HFYEQNERIANGSVNGQLNFNSLGIINGIIDAIOAPYPERAVNNYTIKAVNETVYN 282
216 LIHSLN-----PVREKVINLNGIAGDYSDPESIIGGYAEF-----LYQIGLLDEKQK 265
283 YMKFANQMPNGCDLITCKQTN---RTALADYALCAEATN---MCRDNVSG--PYAFAG 335
266 Y--FQKO---CHCIEHRIKQNFWEAFELDKLLDGLTSDPSYFONVIGCSNYNFL- 318
336 RGVDYDIRHPYDDPPPPSYNKKFLAKDSVMDAIGVININTOSNNDVY--YAFQOQGFDFWVP 393
319 -----RCTEPEDQLYYVKFLSLPEVRQAIHVG-NQTFNDGTIVEKYLREDTVQSVKP 369
394 NFIEDLEEILALPVRSVLIYGDADYI-----CNWFGQAVSLAANYSOAAQ-- 439
370 WLTEIMNN-----YKVLIIYNGQLDIIVAAALTEKSLMGMMDWKSGO-----EYKKAERKV 418
440 ---FRS---AGYTPLKVNVEYGETREYGNFSTRVVEAGHEVPPYQPIASLQIFNRTI 492
419 WKIFKDSSEVAGY-----IRQAGDFHOVIINGGGHILPYDQPLRAFDWINEFI 466
493 F--GWD 496
467 YGKGWD 472

RESULT 13
US-09-907-613-164
; Sequence 164, Application US/09907613
; Publication No. US20030027145A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fond, Sherman

; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: 10466-14
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,613
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-613-164

Query Match      12.4%; Score 359; DB 11; Length 476;
Best Local Similarity 26.7%; Pred. No. 2.3e-25;
Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;

47 PG--VKSYSGY--VDTSPESTHTFFWFEARHNPETAPITLWLNKGPGSDSLICLFEELGP 102
   || :|||:| : : : :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
67 PGLNMKSYAGFLTVNKTYNSNLFWFPPAQIQPEDAPVVLWLOGGPGGSSMFLGFVEHGP 126

103 CHVNSTFDDYINPHSNWNEVSNLLFSLQPLGVGFSDIVDGSINPVTGVVENSFAQVQG 162
127 YVVTSMNLTDRDRFPWTTTSLMLYIDNPVGTGSFTDDTHG----- 167
163 RYPTIDATLDTINLAAEAAWEILOGLFSLGSLDSRVQSKDFSLWFSYGGHYGAFFN 222
168 -----YAVNEDDVARDLYSALIQFE-----QIFPEYKKNDDYVVTGESYAGYVPAIAH 215
223 HFYEQNERIANGSVNGQLNFNSLGIINGIIDAIOAPYPERAVNNYTIKAVNETVYN 282
216 LIHSLN-----PVREKVINLNGIAGDYSDPESIIGGYAEF-----LYQIGLLDEKQK 265
283 YMKFANQMPNGCDLITCKQTN---RTALADYALCAEATN---MCRDNVSG--PYAFAG 335
266 Y--FQKO---CHCIEHRIKQNFWEAFELDKLLDGLTSDPSYFONVIGCSNYNFL- 318
336 RGVDYDIRHPYDDPPPPSYNKKFLAKDSVMDAIGVININTOSNNDVY--YAFQOQGFDFWVP 393
319 -----RCTEPEDQLYYVKFLSLPEVRQAIHVG-NQTFNDGTIVEKYLREDTVQSVKP 369
394 NFIEDLEEILALPVRSVLIYGDADYI-----CNWFGQAVSLAANYSOAAQ-- 439
370 WLTEIMNN-----YKVLIIYNGQLDIIVAAALTEKSLMGMMDWKSGO-----EYKKAERKV 418
440 ---FRS---AGYTPLKVNVEYGETREYGNFSTRVVEAGHEVPPYQPIASLQIFNRTI 492
419 WKIFKDSSEVAGY-----IRQAGDFHOVIINGGGHILPYDQPLRAFDWINEFI 466
493 F--GWD 496
467 YGKGWD 472

RESULT 13
US-09-907-613-164
; Sequence 164, Application US/09907613
; Publication No. US20030027145A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fond, Sherman
```



Db 127 YVTSNMTLRDRDPFWTTLSMLYIDNPVGTGFSFTDDTHG----- 167  
QY 163 RYPTIDATLIDTTNLAAEAAWEILQGLSLGSLDSRVQSKDFSLWTSYSGHYGPAPFN 222  
Db 168 -----YAVNEDDVARLYSALIQFF-----QIFPEYKNDFFYVTGESYAGKYVPAIAH 215  
QY 223 HFYEQNERIANGSVNGVOLNENSLGIINGIIDEAIOAPYPEFVANNNTYGIKAVNETVYN 282  
Db 216 LIHSLN-----PVREVKINLNGIAGDGYSDPESIIGGYAEF-----LYQIGLLDEKQKK 265  
QY 283 YMKFANOMPNGCODLISTCKQTN---RTALADYALCAEATN---MCRDNVEG---PYAFAG 335  
Db 266 Y--FQKQ-----CHECHIRKQNWFEAFELDKLIDGLTSDPSYFQNVGTGCSNYNLF- 318  
QY 336 RGVYDIRHPYDDPTPPSYNKNFLAKDSYMDAIGVNIYNTOSNNDVY---YAFQOTGDFVWP 393  
Db 319 -----RCTEPEQDLYYKFLSLPEVRQAIHVG-NQTFNDGTVIEXYLEDIVQSVKP 369  
QY 394 NFIEDLEIILALPVRVSLIYGADYI-----CNWFGQAVSLAANYSQAAQ-- 439  
Db 370 WLTEIMNN-----YKVLIIYNGQLDIIVAAALTERSLMGMDWKGSQ-----EYKKAQKV 418  
QY 440 ---FRS-----AGYTPLKVNQVEYGETREYGNFSTRVYEAGHEVPYQPIASLOLFNRTI 492  
Db 419 WKLFKSDSEVAGY-----IROAGDFHQVILIRGGGHILPYDQPLRAFDMINRFI 466  
QY 493 F--GWD 496  
Db 467 YKKGWD 472

## RESULT 14

US-09-907-942-164  
Sequence 164, Application US/09907942  
Publication No. US20030027146A1

GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, A.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kljavin, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: 10466-14  
CURRENT APPLICATION NUMBER: US/09/907,942  
CURRENT FILING DATE: 2002-01-22  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222

PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: PCT/US99/28214  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30  
PRIOR APPLICATION NUMBER: PCT/US99/28564  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/28565  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30095  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: PCT/US99/30911  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US99/30999  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US00/00219  
PRIOR FILING DATE: 2000-01-05  
NUMBER OF SEQ ID NOS: 423  
SEQ ID NO 164  
LENGTH: 476  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-907-942-164

Query Match 12.4%, Score 359; DB 11; Length 476;  
Best Local Similarity 26.7%, Pred. No. 2.3e-25;  
Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;

QY 47 PG--VKSYGY--VDISPESHTEFFEFARINPETAPITWLNQPGSDSLIGLFEELGP 102  
Db 67 PGLNMKSYAGFLTIVKNTYNSLNFWEFFPAQIQEDAPVVLWLOQPGSGSMFGLVEHGP 126  
QY 103 CHVNSTEDDYINPHSWNEVSNLLFLSOPLGVPFSYSDTVGSGINPVTGWENSSFAGVQG 162  
Db 127 YVTSNMTLRDRDPFWTTLSMLYIDNPVGTGFSFTDDTHG----- 167  
QY 163 RYPTIDATLIDTTNLAAEAAWEILQGLSLGSLDSRVQSKDFSLWTSYSGHYGPAPFN 222  
Db 168 -----YAVNEDDVARLYSALIQFF-----QIFPEYKNDFFYVTGESYAGKYVPAIAH 215  
QY 223 HFYEQNERIANGSVNGVOLNENSLGIINGIIDEAIOAPYPEFVANNNTYGIKAVNETVYN 282  
Db 216 LIHSLN-----PVREVKINLNGIAGDGYSDPESIIGGYAEF-----LYQIGLLDEKQKK 265  
QY 283 YMKFANOMPNGCODLISTCKQTN---RTALADYALCAEATN---MCRDNVEG---PYAFAG 335  
Db 266 Y--FQKQ-----CHECHIRKQNWFEAFELDKLIDGLTSDPSYFQNVGTGCSNYNLF- 318  
QY 336 RGVYDIRHPYDDPTPPSYNKNFLAKDSYMDAIGVNIYNTOSNNDVY---YAFQOTGDFVWP 393  
Db 319 -----RCTEPEQDLYYKFLSLPEVRQAIHVG-NQTFNDGTVIEXYLEDIVQSVKP 369  
QY 394 NFIEDLEIILALPVRVSLIYGADYI-----CNWFGQAVSLAANYSQAAQ-- 439  
Db 370 WLTEIMNN-----YKVLIIYNGQLDIIVAAALTERSLMGMDWKGSQ-----EYKKAQKV 418  
QY 440 ---FRS-----AGYTPLKVNQVEYGETREYGNFSTRVYEAGHEVPYQPIASLOLFNRTI 492  
Db 419 WKLFKSDSEVAGY-----IROAGDFHQVILIRGGGHILPYDQPLRAFDMINRFI 466  
QY 493 F--GWD 496  
Db 467 YKKGWD 472





Db 383 TVDNFSLKVGAGHEVYQPTALQAFKQII-----QKK 418

RESULT 2

S51516  
serine-type carboxypeptidase (EC 3.4.16.-) 2 precursor - Absidia zycharae  
C:Species: Absidia zycharae  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 16-Jun-2000  
C:Accession: S51516; S78013; S78014  
R:Lee, B.R.; Takeuchi, M.; Kobayashi, Y.  
Curr. Genet. 27, 159-165, 1995  
A:Title: Molecular cloning and sequence analysis of the scp2 gene encoding the serine carboxypeptidase from Absidia zycharae  
A:Reference number: S51516; MUID:95308538; PMID:7788719  
A:Accession: S51516  
A:Molecule type: DNA  
A:Residues: 1-460 <LEE>  
A:Cross-references: EMBL:D16519; NID:g556466; PIDN:BAA03966.1; PID:g995456  
A:Experimental source: strain NRIC 1199  
A:Accession: S78013  
A:Molecule type: protein  
A:Residues: 52-62; 90-99; 367-381 <LEB>  
A:Accession: S78014  
A:Molecule type: mRNA  
A:Residues: 18-460 <LES>  
C:Genetics:  
A:Gene: spcZ  
A:Introns: 64/3; 95/1; 107/2; 138/3; 234/2; 246/1; 270/1; 343/3; 353/2; 384/3; 437/3  
C:Superfamily: serine carboxypeptidase  
C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:18-51/Domain: propeptide #status predicted <PRE>  
F:52-460/Product: serine-type carboxypeptidase #status experimental <MAT>  
F:40,46,119/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:186/Active site: Ser #status predicted

Query Match 25.9%; Score 746.5; DB 1; Length 460;  
Best Local Similarity 35.8%; Pred. No. 3.5e-44;  
Matches 175; Conservative 71; Mismatches 182; Indels 61; Gaps 14;

QY 6 PASVGRQLPKNTGVKTLTANNVIRYKEPAGVCEVTPGVKSYGVVDSPESHF 65  
Db 29 PMSHVLRRQGNPT-----SSGNTLTQLKTSF---KLCD--PDVKQSYGLVDANDSHYF 78

QY 66 FWFEEARHPNPETAPITLWNGGPGSDSLIGLFEELGPGCHVNSFTDDYINPHSWNEVSNLL 125  
Db 79 FWFEEKNDKPNPLIWLNGGPGSCSLIGLWEELGPGQNGS----ANPHSWHHSSNML 134

QY 126 FLSQPLGVGSYSVDVDSINPVTVGVVENSFAGVQGRYPTIDATLIDTNLAEEAAWEI 185  
Db 135 FFDQPDGVGTSY-----GKQ-----TVSTTTEDAERAATF 164

QY 186 LQGLFSLGPSLDSRVQSKDSLWTESYGGHYGPAFFNFHFEONRIANGSVNGVQLNFS 245  
Db 165 LQAFYETFP---QYSKLDVHYGESYGGHYIPGFASHVVDMMKNKQSGEGKGVVPLKS 220

QY 246 LGIINGITDAIAQPYPEFVANNNTYGIKAVNETVYNNMFANQMPGQDLSTCKQTN 305  
Db 221 IGVNGFLDAVIOYKSPKTKCHSTYPAVLSE---ECDKMQQIYENDCKPAAEQCAESD 277

QY 306 RTALADYALCAEATNMCRONVEGYPYAFAGVGVYDIHRPYDDPTPPSYNNKFLAKDSMD 365  
Db 278 -----EDSDCVNANQCC-GQTEG-IYAQSGYSFYDIRQQGDD-TPHPFVDR-LNKASVIK 328

QY 366 AIGVNIQNTQSNVDVYAFQGTDFVWPNFTLEELIALLPVRSYLIYGDADYICNNFEGG 425  
Db 329 EVGARGHFSMCSDSVGTAFATGDCA-RSYIPAVEKLLKEGIPVLIVGDDADVVCNWCN 387

QY 426 QAVSLAANYSOAAQFRSAGYTPLVKNGVVEYGETREYGNFSFTRYEAGHEVYQPTIASL 485  
Db 388 LDVADSLKWDGSDAFSKTKLEAWKADGKEVQGRSADKLTEVRVYEAGHEVPMYQPEAL 447

QY 486 QLENKNTIFG 494

Db 448 SMFQTWISG 456

RESULT 3

T43236  
carboxypeptidase C (EC 3.4.16.5) precursor [validated] - fission yeast (Schizosaccharomyces pombe)  
N:Alternate names: carboxypeptidase Y  
C:Species: Schizosaccharomyces pombe  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 08-Sep-2000  
C:Accession: T43236; T37997  
R:Tabuchi, M.; Iwaihara, O.; Ohtani, Y.; Onuchi, N.; Sakurai, J.; Morita, T.; Iwahara, J.; Bacteriol. 179, 4175-4189, 1997  
A:Title: Vacuolar protein sorting in fission yeast: cloning, biosynthesis, transport, and localization of the vacuolar carboxypeptidase Y  
A:Reference number: T22357; MUID:97352672; PMID:9209031  
A:Accession: T43236  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1002 <TAB>  
A:Cross-references: EMBL:D86560; NID:g3046860; PIDN:BAA25568.1; PID:g3046861  
R:Oliver, K.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
Submitted to the EMBL Data Library, June 1997  
A:Reference number: Z21760  
A:Accession: T37997  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1002 <OLI>  
A:Cross-references: EMBL:Z21760; PIDN:CABL0121.1; GSPDB:GM00066; SPDB:SPACL9GL2.10C  
A:Experimental source: strain 972h(-); cosmid c19GL2  
C:Comment: The mature carboxypeptidase exists as a heterodimer held together by a disulfide bond.  
C:Genetics:  
A:Gene: cpy1; SPACL9GL2.10C  
A:Map position: 1  
C:Superfamily: serine carboxypeptidase  
C:Keywords: disulfide bond; glycoprotein; heterodimer; hydrolase; serine carboxypeptidase  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-521/Domain: propeptide #status predicted <PRO>  
F:200-332/Region: 13-residue repeats (H-H-[EK]-p-G-E-H-M-p-p-p-[MF])  
F:344-425/Region: 9-residue repeats (E-H-H-[RKQ]-G-G-[KE]-[DE]-K)  
F:522-1002/Product: carboxypeptidase C #status experimental <MAT>  
F:627-880/Disulfide bonds: #status predicted  
F:659/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:715,921,978/Active site: Ser, Asp, His #status predicted

Query Match 22.6%; Score 652.5; DB 2; Length 1002;  
Best Local Similarity 33.8%; Pred. No. 3.4e-37;  
Matches 165; Conservative 66; Mismatches 174; Indels 83; Gaps 15;

QY 31 TIRYKEPAGVCEVTPGVKSYGVVDSPESHFFWFEEARHPNPETAPITLWNGGPGS 90  
Db 570 TLRKDSKPSGLGDT--VKQYTGVLVDVDRHLEFFWFEESENDEPNPVLWNGGPGC 627

QY 91 DSLIGLFEELGPGCHVN-STFDDYINPHSWNEVSNLLFSLQPLGVGSYSVDVDSINPVT 149  
Db 628 SSLTGLEMLGELPSSINTELKPEYNPHSWNSNASVIFLDQPTINGFSGND--DSVLDIVT 685

QY 150 GVWENSFAGVQGRYPTIDATLIDTNLAEEAAWEILQGLSPLDSRVQSKDFSLWT 209  
Db 686 -----AGKDVYAEALNLFKAFP-----QTAHLDFHAG 713

QY 210 ESYGHHYGPAPFNHFEONER-----IANG-SVNGVQLNFSLIGITINGIDEAIAQPYYPE 264  
Db 714 ESYAGHYIYQPAKETMEHNOQANFVSGYEMEKQYINLKSVLINGLTDPLVQYFEGK 773

QY 265 FAVNNYTG-----IKAVNETVYNNMFANQMPGQDLSTCKQTNRIALADYAL 314  
Db 774 MACSPYGLIMSQEECDRIIGAYDT-----CAKLITGQYGTGFI-----PV 814

QY 315 CAEATNCRDNVEGYPYAFAGVGVYDIHRPYDDPTPPSY-----YNNFLAKDSVMDATGV 369  
Db 815 CIGASLCNNAMIGP-FTKTGINIYDIREECDQHEHCYPTGATESILNQEFVEALCV 873

QY 370 NINYTQSNVDVYAFQGTDFVWPNFTLEELIL--ALPVRVSLYI-GDADYICNWFEGG 426

Db 874 EYDYGKGVNTEVNFELFKGDMRKTRDDVIAILEAGLPV---LIYAGDADYICNMGNE 930  
 QY 427 AVSLAANYSQAAQFRSAGYTPLVKNGVEYGETREVGNSFTRVYAGHEVYQIASIQ 486  
 Db 931 AWTDALEWAGQEFYEAEKFWSPNGKAGRGKSKFNFGYLRKLYEAGHVPFNOPEASLE 990  
 QY 487 LENRTIFG 494  
 Db 991 MLNSWIDG 998

## RESULT 4

S46008

Probable carboxypeptidase C (EC 3.4.16.5) YBR139w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YBR1015; hypothetical protein YBR139w

C:Species: Saccharomyces cerevisiae

C:Date: 26-Aug-1994 #sequence\_revision 05-Jan-1996 #text\_change 19-Apr-2002

C:Accession: S46008; S46581

R:Becam, A.M.; Herbert, C.J.; Nasr, F.; Slonimski, P.P.; Zagulski, M.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45995

A:Accession: S46008

A:Molecule type: DNA

A:Residues: 1-508 <BEC>

A:Cross-references: EMBL:Z36008; NID:G536435; PIDN:CAA85097.1; PID:G536436; GSPDB:GN0000

A:Experimental source: strain S288C

R:Becam, A.M.; Cullin, C.; Grzybowski, E.; Lacroute, F.; Nasr, F.; Ozier-Kalogeropoulos,

Yeast 10(Suppl.A), S1-S11, 1994

A:Title: The sequence of 29.7kb from the right arm of chromosome II reveals 13 complete

A:Reference number: S46569; MUID:94378717; PMID:8091856

A:Accession: S46581

A:Molecule type: DNA

A:Residues: 1-508 <BE2>

A:Cross-references: EMBL:X75891; NID:G496856; PIDN:CAA53497.1; PID:G496869

A:Experimental source: strain S288C

C:Genetics:

A:Cross-references: SGD:S0000343

A:Map position: 2R

A:Note: MIPS:YBR139w

C:Superfamily: serine carboxypeptidase

C:Keywords: hydrolase; serine carboxypeptidase; transmembrane protein

F:6-22/Domain: transmembrane #status predicted <TM>

F:219,415,474/Active site: Ser, Asp, His #status predicted

Query Match 22.0%; Score 636; DB 1; Length 508;  
 Best Local Similarity 32.6%; Pred. No. 1.9e-36;  
 Matches 167; Conservative 70; Mismatches 197; Indels 78; Gaps 14;

QY 1 LPGSTPASVGRRL-----PKNPFGVKLTITANNVTTRYKEPGAEGVCTTPGVKSYGYV 56  
 Db 43 LFQNTQOTLKLRNHHDDPLFTTFISSVDITYSLRLTVDPSKLGIT---DTVKQSGYM 98  
 QY 57 DTPSESTFFWFPEARHPNPETAPITLWLGPGSDSIIGLFGELGCHVNSTEDDYINPH 116  
 Db 99 DYKDSKHFFYWFESRNDPNDPIILWLGPGCCSFTGLLFGELGCHVNSTEDDYINPH 158  
 QY 117 SNNVSNLLFLSQPLGVGSYSDTVDSINPVTGVVENSFAGVQGRYPTIDATLIDTIN 176  
 Db 159 SNNVSNMIFLEQPLGVGSYSD-----EKVSSYK 188  
 QY 177 LAARAAWEILOGLSLPSLDSRVSKDPSLWTEYGGHYGPAFFNHFEQN-ERIANGS 235  
 Db 189 LAGKDAYIFLELFEAFPHL-----RSNDPHIAGESYAGHYIPIQIAHEIVVKNER----- 239  
 QY 236 VNGVQLNFSNLGTINGLIDEALQAPVYPEFV-NNTYGIKAVNEIVNTYMKFANQMPNCG 294  
 Db 240 -----TFNLTSVMIGNITDPLQADYEPMACGKGHYEVLSSSECEKMSKAAGR----- 290  
 QY 295 ODLSCTKQTNRIALADYAL--CAEATNCRONVEGPGYVAFAGRGVYDIRHYPDDP----- 348  
 Db 291 -----CRRLNLCYASKSSLPCIVATYACDSALLEPYIN-TGLNVYDIRGCEDNSTDG 343  
 QY 349 ---TPPSYNKFLAKDSVDAIGVNI-NYTSQNNVDVYAFQOTGDFVWPNFTEDLEEILA 404

Db 344 MCYTGLRVDDQZMNFPEVQETLGSVDVHNYSGCDNVFTGLFTGDSKPK-FQYIAELLN 402  
 QY 405 LPVRSLYIGDADYICNNMFGGCAVSLAANYSQAAQFRSAGYTP--LKVNGVEYGETREVG 462  
 Db 403 HNPVLIYAGDKDYICNWLGNHANSNELEWINKRYQRRMLRPWWSKETGEELGOVKNYG 462  
 QY 463 NESFTRVYAGHEVYQIPIASLOLNFRTIFG 494  
 Db 463 PFTFLRIYDAGHMVPYDQPEASLEMVNSWISG 494

## RESULT 5

JC7666

Serine-type carboxypeptidase homolog precursor - Emericella nidulans

N:Alternate names: carboxypeptidase Y homolog

C:Species: Emericella nidulans, Aspergillus nidulans

C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001

C:Accession: JC7666

R:Ohsumi, K.; Matsuda, Y.; Nakajima, H.; Kitamoto, K.

Biosci. Biotechnol. Biochem. 65, 1175-1180, 2001

A:Title: Cloning and characterization of the cpvA gene encoding intracellular carboxy

A:Reference number: JC7666; MUID:21333188; PMID:11440134

A:Accession: JC7666

A:Molecule type: mRNA

A:Residues: 1-552 <OHS>

A:Cross-references: DDBJ:AB051820

C:Comment: This protein is an intracellular carboxypeptidase, which is localized in v

C:Genetics:

A:Gene: cpvA

A:Introns: 165/1

F:1-15/Domain: signal sequence (or 1-18, or 1-19) #status predicted <SIG>

F:16-131/Domain: pro-sequence, vacuolar targeting signal (or 19-131, or 20-131) #stat

F:131-132/Region: cleavage site, by Arg-11e #status predicted

Query Match 20.8%; Score 601; DB 2; Length 552;  
 Best Local Similarity 31.6%; Pred. No. 5.6e-34;  
 Matches 154; Conservative 77; Mismatches 175; Indels 82; Gaps 19;

QY 27 ANNVTIRYKPGAEGVCTTPGVKSYGYVDTSP-ESHTFFWFPEARHPNPETAPITLWLN 85  
 Db 127 AYDLRIKKTDPSSLGI--DPVKQYIGYLLDNDKHLFWFESRNDPANDPVVLWLN 183  
 QY 86 GPGSDSLIGLFGELGCHVNSTFDDYINPHSNVSNLLFLSQPLGVGSYSDTVDSGI 145  
 Db 184 GPGCCSLTLGLFELGELGCHVNSTFDDYINPHSNVSNLLFLSQPLGVGSYSDTVDSGI 205  
 QY 146 NPVTGVVENSFAGVQGRYPTIDAILLDTNLAARAAWEILOGLSLPSLDSRVQSKDF 269  
 Db 237 -----GSTVSDIV-AAGKDYALLTLFFKQFP-----EYARQDF 269  
 QY 206 SLWTEYGGHYGPAFFNHFEQNRIANGSVNGVQLNFSNLGTINGLIDEALQAPVYPEF 265  
 Db 270 HIAGESYAGHYIPVFTSEILSHQK-----NINLKSVLINGLTLGTLQVYETPRM 320  
 QY 266 AVNNTYGIKAVNEIVNTYMKFANQMPN-----ODLSCTKQTNRIALADYALCAEATNMC 322  
 Db 321 ACCEG-GYPAVLDE-----SSCRSDNALGRCSMTESCYNSESA-----WVCVPASIYC 369  
 QY 323 RONVEGPIYAFAGRGVYDIRHYPDDPT-----PPSYNKFELAKDSVDAIGVNI-YTQS 376  
 Db 370 NNALLAPYOR-TGQNYDVRGKCEDESNCYKGMGVSEYLNKPEVRAAAGVSDGYDSC 428  
 QY 377 NNDVYFAQOTGDFVWPN--NFTEDLEEILALPVRSLLY-GDADYICNNMFGGCAVSLAAN 433  
 Db 429 NFDINNFLFHGDWKKPYHRLVPGILE--QIPV---LIYAGDADFTCNLGNKAWTEALE 483  
 QY 434 YSQAQAFRSAGYTPLVK-----NGVEYGETREYGNFSFTRVYAGHEVYQIPIASLOLF 488  
 Db 484 WFGHKEFAAAPMEDLKIVDNEHTGKIGQIKTHGNTFMRLYGGGHHVWMDQPEASLEFF 543  
 QY 489 NRTIFG-W 495  
 || : : |



[illegible][illegible]

Db 379 SYGSLFLKLVHDAGHVPMDQPKVALEMLRWTSQ 413

RESULT 10  
S22530  
carboxypeptidase C (EC 3.4.16.5) precursor - rice  
N:Alternate names: carboxypeptidase III  
C:Species: Oryza sativa (rice)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 24-Nov-1999  
C:Accession: S22530  
R:Washio, K.; Ishikawa, K.  
Plant Mol. Biol. 19, 631-640, 1992  
A:Title: Structure and expression during the germination of rice seeds of the gene for a  
A:Reference number: S22530; MUID:92329723; PMID:1627776  
A:Accession: S22530  
A:Molecule type: DNA  
A:Residues: 1-500 <WAS>  
A:Cross-references: EMBL:S40458  
C:Genetics:  
A:Introns: 100/2; 156/3; 196/3; 244/3; 295/1; 324/3; 411/1; 455/3  
C:Superfamily: serine carboxypeptidase  
C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase; zymogen  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-73/Domain: propeptide #status predicted <PRO>  
F:74-484/Product: carboxypeptidase C #status predicted <MAT>  
F:485-499/Domain: carboxyl-terminal propeptide #status predicted <CPRO>  
F:144/Binding site: carboxylate (Asp) (covalent) #status predicted  
F:216,404,461/Active site: Ser, Asp, His #status predicted

Query Match 15.7%; Score 452; DB 1; Length 500;  
Best Local Similarity 26.4%; Pred. No. 1.1e-23;  
Matches 128; Conservative 81; Mismatches 214; Indels 62; Gaps 14;  
QY 14 LPK--NPTGVKTLTANNVIRKEPCAGVCETPGVSKYSGY--VDTSPESTHFWFF 69  
DB 47 LPKAGPTAGDVPSPVAPGELLERRVTLPLPGQVGLGHAGYRPLNTHDARMTFLF 106  
QY 70 EARHNPETAPITLWNGPGSDSLIGLFEELGPGHVNSTFDDYINPHSNWNSNLLFLSQ 129  
DB 107 ESRGKKED-PVWILTGPGCSSELAVFYENGPTTISNNKSLANWKGWDFTISNIFVDQ 165  
QY 130 PLGVGFYSYDVTGGINPVTGVVNSFSFAGVQGRYPIDATLITDNLAAEAWEILQCF 189  
DB 166 PTGTGFYSYSDDRDTRHDTGV-----SNDLYSFLQVF 198  
QY 190 LSGPLSDRSVQKDFSLWTFESGHHYGPAPFNHFYENONRTANGSVNGVQNLNLSLII 249  
DB 199 FKKHPEF-----AKNDFITGESYAGHYIPAFSRVHQNK--AN---EGIHNLKGFAIG 249  
QY 250 NGIIDEATQAPYEPFAVNTYGIKAVNETVYNTMKFANQMPNGCQDLISTCKQTNRTAL 309  
DB 250 NGLTDPALQYKAYTDYALD-----MNLIKKSDYDRI---NKFIPCEFAIKLCGTNGK--- 299  
QY 310 ADYALCAEATNMCRDNVEGPPYAFAGRVYDRIHPYDPTPPSYN--KFLAKDSYMDAI 367  
DB 300 ---ASCAAAVMVCNLSIFSSIMKLVGTKNKYDVRKECEGKLCYDFSNLEKFTGDKAVKEAL 356  
QY 368 GV-NINYTOSNNDVYAFQOQGDFFWPNFIE-DLEEIILALPVRVSLIYGDADYICNWFEGG 425  
DB 357 GVGDLFEVSCSTTVQAMLD-----WMENLEVGIPALLEDGINVLIIYAGEYDLICNWLGN 412  
QY 426 QAVSLAANYSOAOFRSAGYTLKYNVEYGETREYGNFSTFRVYEAAGHEVYYPPIASL 485  
DB 413 SRWHISMWMSGKDKDFVSSHESPEFVVDGAEAGVLKSHGSLPLKVNHAGHVMVMDQPKASL 472  
QY 486 QLFNR 490  
DB 473 EMLRR 477

RESULT 11  
T48977  
carboxypeptidase-like protein F14D17.80 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 24-Oct-2000  
C:Accession: T48977  
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z25008  
A:Accession: T48977  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-510 <JOR>  
A:Cross-references: EMBL:AL353992; GSPDB:GN000061; ATSP:F14D17.80  
A:Experimental source: cultivar Columbia; BAC clone F14D17  
C:Genetics:  
A:Gene: ATSP:F14D17.80  
A:Map position: 3  
A:Introns: 115/2; 171/3; 259/3; 310/1; 341/3; 428/1; 472/3  
C:Superfamily: serine carboxypeptidase  
F:231,421,478/Active site: Ser, Asp, His #status predicted

Query Match 15.2%; Score 439; DB 2; Length 510;  
Best Local Similarity 26.6%; Pred. No. 8.8e-23;  
Matches 123; Conservative 81; Mismatches 189; Indels 69; Gaps 14;  
QY 46 TPGVKSYS--SGY--VDTSPESTHFWFFEARHNPETAPITLWNGPGSDSLIGLFEEL 100  
DB 93 SPSVQDFGHARGYKLPNSKAARMFYFFESRTN-KADPVVILWLTGGPGCSSELALFVEN 151  
QY 101 GPCHVNSTFDDYINPHSNWNSNLLFLSQPLGVGFYSYDVTGGINPVTGVVNSFSFAGV 160  
DB 152 GPTVSNNSSLSWNEFGWDKASNLIIYDQPVGTGFSYTSQSDLRHDEHDGV----- 202  
QY 161 QGRYPTIDAILDITNLAAEAWEILQGLSLGSLDSRVQSKDFSLWTFESGHHYGPAPF 220  
DB 203 -----SNDLYDLQAFKHEP-----QFVKNDIFYITGESYAGHYIPAL 240  
QY 221 FNHEYEONERTANGSVNGVQNLNLSLIGIIDEAIOAPYEPFAVNTYGIKAVNETV 280  
DB 241 ASRVHRGNK-----NKEGTHNLKGFAIGNLNPEIOYGAADYALDNLITQSDHDL 295  
QY 281 YNYMKFANQMPNGCQDLISTCKQTNRTALADYA---LCAEATNMCRDNVEGPPYAFAGR- 336  
DB 296 NRY-----YATQOQSTKECSADGEGDACCASYTVC--NNIFOKIMDIAGNV 340  
QY 337 GYDTRHPYDPTPPSYN--KFLAKDSYMDAIGV-NINYTOSNNDVYAFQOQGDFFWVP 393  
DB 341 NYDVRKQCEGLCVDFSNMENFLNOKSVKALGVGDIEFVSCSTAVTEAMQMD---WM 396  
QY 394 NFIE-DLEEIILALPVRVSLIYGDADYICNWFEGGOAVSLAANYSOAOFRSAGYTLKYN 452  
DB 397 RNLEVGIPALLEDGINVLIIYAGEYDLICNWLGNKSWHMEWMSGKKEFVAAATVEFHDN 456  
QY 453 VEYGETREYGNFSTFRVYEAAGHEVYYPPIASLQLFNRTIFG 494  
DB 457 KEAGLMKNYGSFLTKLVHDAGHVPMDQPKAALQMLQNMVQ 498

RESULT 12  
A35275  
carboxypeptidase C (EC 3.4.16.5) - barley  
N:Alternate names: carboxypeptidase III  
C:Species: Hordeum vulgare (barley)  
C:Date: 24-Nov-1999 #sequence\_revision 24-Nov-1999 #text\_change 24-Nov-1999  
C:Accession: A35275  
R:Sorensen, S.B.; Svendsen, I.; Bredam, K.  
Carlsberg Res. Commun. 54, 193-202, 1989  
A:Title: Primary structure of carboxypeptidase III from malted barley.  
A:Reference number: M35275; MUID:90315015; PMID:2639682  
A:Accession: A35275  
A:Molecule type: protein  
A:Residues: 1-411 <SOR>  
A:Note: 185-Val was also found  
C:Superfamily: serine carboxypeptidase  
C:Keywords: blocked amino end; glycoprotein; hydrolase; serine carboxypeptidase







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 22:35:07 ; Search time 33 seconds  
(without alignments)  
765.253 Million cell updates/sec

Title: US-09-712-338-2\_COPY\_19\_555  
Perfect score: 2887  
Sequence: 1 LPSGTPASVGRRLPKNPTG.....HTQSSVPLPTATSMSSVGM 537

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	894	31.0	423	1	CPSL_PENJA
2	652.5	22.6	1002	1	CBPY_SCHPO
3	636	22.0	508	1	YBY9_YEAST
4	599	20.7	542	1	CBPY_CANAL
5	594.5	20.6	523	1	CBPY_PICPA
6	527.5	18.3	532	1	CBPY_YEAST
7	474	16.4	516	1	CBPX_ARATH
8	473	16.4	429	1	CBPX_ORISA
9	452	15.7	500	1	CBP3_ORISA
10	434	15.0	508	1	CBP3_HORVU
11	413.5	14.3	500	1	CBP3_WHEAT
12	359	12.4	476	1	CPVL_HUMAN
13	353.5	12.2	436	1	CP22_HORVU
14	352.5	12.2	470	1	YSS2_CAEEL
15	351	12.2	482	1	NF31_NAEFO
16	349	12.1	474	1	PRTP_MOUSE
17	344.5	11.9	516	1	YUW5_CAEEL
18	332	11.5	469	1	CP23_HORVU
19	328.5	11.4	480	1	PRTP_HUMAN
20	326.5	11.3	454	1	YUA6_CAEEL
21	321.5	11.1	729	1	KEX1_YEAST
22	318.5	11.0	476	1	CBP2_HORVU
23	316	10.9	574	1	YVP3_CAEEL
24	312.5	10.8	471	1	VCP_AEDAE
25	312	10.8	510	1	CBPL_ORISA
26	305.5	10.6	423	1	CBP2_WHEAT
27	297.5	10.3	505	1	YXD2_CAEEL
28	295.5	10.2	286	1	CBPX_PEA
29	280	9.7	499	1	CBPL_HORVU
30	269.5	9.3	523	1	PEPS_ASPSA
31	255	8.8	452	1	RISC_MOUSE
32	248	8.6	452	1	RISC_RAT
33	237	8.2	452	1	RISC_HUMAN

34	221.5	7.7	531	1	PEPF_ASPNG
35	219.5	7.6	507	1	SXA2_SCHPO
36	167.5	5.8	324	1	CP21_HORVU
37	147	5.1	366	1	HNLS_SORBI
38	120.5	4.2	1061	1	OAR_MYXXA
39	114.5	4.0	717	1	AMVM_BACST
40	113.5	3.9	1287	1	RPOL_FOWPV
41	112	3.9	376	1	OE56_NPVAC
42	112	3.9	787	1	AGL2_BACTQ
43	111	3.8	696	1	OXLA_NEUCR
44	110.5	3.8	1196	1	AMVB_PAEPO
45	109.5	3.8	481	1	PRTB_ERWCH

ALIGNMENTS

RESULT 1					
CPSL_PENJA					
ID	CPSL_PENJA	STANDARD;	PRI;	423 AA.	
AC	P34946;				
DI	01-FEB-1994 (Rel. 28, Created)				
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DE	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Carboxypeptidase S1 (EC 3.4.16.6).				
OS	Penicillium janthinellum (Penicillium vitale).				
CC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;				
CC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.				
OX	NCBI_TaxID=5079;				
RN	[1]				
RP	SEQUENCE.				
EX	MEDLINE=94039747; PubMed=8224168;				
RA	Svendsen I., Hofmann T., Endrizzi J., Remington S.J., Breddam K.;				
RI	"The primary structure of carboxypeptidase S1 from Penicillium				
RT	janthinellum";				
RL	FEBS Lett. 333:39-43(1993).				
CC	-!- CATALYTIC ACTIVITY: Preferential release of a C-terminal arginine				
CC	or lysine residue.				
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.				
DR	PIR; S38953; S38953.				
DR	MEROPS; S10.008; -.				
DR	InterPro; IPR000379; Ser_estrs_site.				
DR	InterPro; IPR001563; Serine_carbpept.				
DR	Pran; PF00450; serine_carbpept; 1.				
DR	PRINTS; PR00724; CRBOXYPTASEC.				
DR	PRODOM; PD001189; Serine_carbpept; 1.				
DR	PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.				
DR	PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.				
KW	Hydrolase; Carboxypeptidase; Glycoprotein.				
FT	ACT_SITE 143 143				
FT	ACT_SITE 340 340				
FT	ACT_SITE 397 397				
FT	BINDING 343 343				
FT	BINDING 398 398				
FT	DISULFID 8 68				
FT	DISULFID 55 300				
FT	DISULFID 223 246				
FT	DISULFID 230 239				
FT	CARBOHYD 200 200				
SQ	SEQUENCE 423 AA: 46496 MW, 55F3333608B12729 CRC64;				
	N-LINKED (GLCNAC...).				
Query Match	31.0%; Score 894; DB 1; Length 423;				
Best Local Similarity	40.1%; Pred. No. 1.5e-55;				
Matches 186; Conservative	66; Mismatches 160; Indels 52; Gaps 8;				
QY	41 GVCETTPGVKSYGYDTSPESTFFWFEEARNINPETAPITLWLNKGPSSDSLIGLFEEL 100				
Db	6 GICETTPGVKSYGYDTSPESTFFWFEEARNINPETAPITLWLNKGPSSDSLIGLFEEL 100				
QY	101 GPCH-VNSIFDDYINPHSWNEVSNLLFSLQPLGVGFSYDVTGDSINFTVGVVENSFAG 159				
Db	66 GPCHVNSIFDDYINPHSWNEVSNLLFSLQPLGVGFSYDVTGDSINFTVGVVENSFAG 159				

QY 160 VQGRYPTDITADLTNTNLAEEAAWEILQGLSLGSRVQSKDFSLWTSYSGGHGPA 219  
 Db 105 -----TDDVT---STVTAPYVWNLQAFYAPRPEYER---DRAIFTESYGGHGP 151  
 QY 220 FNFHFEYQNRKANGSVNGVQLNFSNLGIINGIIDEAIOAPYYPEFAVNNYIGKAVNET 279  
 Db 152 FASYIEQNAAKAGSVTGNVVALGVNNGWIDSTIQEKAVIDFSYNNYSQOIIDSDS 211  
 QY 280 VYNYMFANPMPCQDLISTCQTNRKTLADYALCAEATNMCRDNVGEGPYAFAGGVY 339  
 Db 212 RDSLLDAYN---NQCLPALQCCSQSGST-----SDCTNADSVCYQIEGPISSGGPDVY 263  
 QY 340 DIRHPYDDPPEPSYINKFKADSGVMDAIGVNTQSNNDVYFAQQTQGFVWPNFTEDL 399  
 Db 264 DIRPSNDPPPKTYSTYLSIDPTVVKALGARTNYQPCPNPKYKFASTGDNPSRSLSTL 322  
 QY 400 EEILALPVRVSLYIGDADYICNFMFGQAVSLAANYSOAQAQFRSAGYTPKLKVNVEYGETR 459  
 Db 323 SSVQSGINVLWAGDADWICNWLGNVEVANAVDFPGNAQFSALDLAPYTVNGVEKGQFK 382  
 QY 460 EYGNFSFTRYEAGHEVPYQPTASLQFLNRTIFGWDIAEGQK 503  
 Db 383 TVDNFSLKYGAGHEVPYQPTALQAFKQII-----QKK 418

RESULT 2  
 CBPY\_SCHPO STANDARD; PRT: 1002 AA.  
 ID CBPY\_SCHPO  
 AC 013849; 014366;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Carboxypeptidase Y precursor (EC 3.4.16.5) (CYP).  
 GN CPY1 OR CPY1L OR SPAC19G12.10C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID:4896;  
 RN [1]  
 RP SEQUENCE FROM N.A., SUBUNIT, SUBCELLULAR LOCATION, AND MUTAGENESIS OF  
 RP CYS-627.  
 RX MEDLINE=97352672; PubMed=9209031;  
 RA Tabuchi M., Iwahara O., Ohtani Y., Ohuchi N., Sakurai J.-I.,  
 RA Morita T., Iwahara S., Takegawa K.;  
 RA "Vacuolar protein sorting in fission yeast: cloning, biosynthesis,  
 RT transport, and processing of carboxypeptidase Y from  
 RT Schizosaccharomyces pombe.";  
 RL J. Bacteriol. 179:4179-4189(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu S., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 Spakovski G.V., Usery D., Bartell B.G., Nurse P.;  
 "The genome sequence of Schizosaccharomyces pombe.";  
 Nature 415:871-880(2002).  
 RL -!- FUNCTION: Involved in degradation of small peptides. Digests  
 preferentially peptides containing an aliphatic or hydrophobic  
 residue in p1' position, as well as methionine, leucine or  
 phenylalanine in p1 position of ester substrate.  
 CC -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a  
 broad specificity.  
 CC -!- SUBUNIT: Heterodimer of two subunits of 32 kDa and 19 kDa derived  
 from the precursor protein and linked by a disulfide bond.  
 CC -!- SUBCELLULAR LOCATION: Lysosome-like vacuoles.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D86560; BAA25568.1; -;  
 DR EMBL: Z97209; CAB10121.1; -;  
 DR PIR: T43236; T43236.  
 DR HSSP: P00729; 1YSC.  
 DR MEROPS: S10.001; -;  
 DR GeneDB\_Spombe: SPAC19G12.10C; -;  
 DR InterPro: IPR000379; Serine\_site.  
 DR InterPro: IPR001563; Serine\_carbpept.  
 DR Pfam: PF00450; serine\_carbpept; 1.  
 DR PRINTS: PR00724; CRBOXYPTASEC.  
 DR PRODOM: P0001189; Serine\_carbpept; 1.  
 DR PROSITE: PS00130; CARBOXYPEPT\_SER\_SER; 1.  
 DR PROSITE: PS00560; CARBOXYPEPT\_SER\_HIS; FALSE\_NEG.  
 DR Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Repeat; Signal.  
 KW SIGNAL 1 18  
 FT PROPEP 19 521  
 FT CHAIN 522 1002  
 FT DOMAIN 208 341  
 FT DOMAIN 225 341  
 FT REPEAT 225 237  
 FT REPEAT 238 250  
 FT REPEAT 251 263  
 FT REPEAT 284 276  
 FT REPEAT 277 289  
 FT REPEAT 290 302  
 FT REPEAT 303 315  
 FT REPEAT 316 328  
 FT REPEAT 329 341  
 FT DOMAIN 361 423  
 FT REPEAT 361 369  
 FT REPEAT 370 378  
 FT REPEAT 379 387  
 FT REPEAT 398 396  
 FT REPEAT 397 405  
 FT REPEAT 406 414  
 FT REPEAT 415 423  
 FT ACT\_SITE 715 715  
 FT ACT\_SITE 921 921  
 FT ACT\_SITE 978 978  
 FT BINDING 924 924  
 FT BINDING 979 979  
 FT DISULFID 627 860  
 FT DISULFID 776 789  
 FT DISULFID 799 822  
 FT DISULFID 806 815





OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=871131100; PubMed=3028649;  
 RA Valls L.A., Hunter C.P., Rothman J.H., Stevens T.H.;  
 RI "Protein sorting in yeast: the localization determinant of yeast  
 RI vacuolar carboxypeptidase Y resides in the propeptide.";  
 RL Cell 48:887-897(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX PubMed=9169872;  
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
 RA Jagels K., Iye G., Moule S., Odell C., Pearson D., Rajandream M.A.,  
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;  
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome  
 RT XIII.";  
 RL Nature 387:90-93(1997).  
 RN [3]  
 RP SEQUENCE OF 112-532.  
 RA Svendsen I., Martin B.M., Viswanatha T., Johansen J.T.;  
 RT "Anionic acid sequence of carboxypeptidase Y. II. Peptides from  
 RT enzymatic cleavages.";  
 RL Carlsberg Res. Commun. 47:15-27(1982).  
 RN [4]  
 RP REVISIONS, AND ACTIVE SITE SER-257.  
 RA Breddam K., Svendsen I.;  
 RT "Identification of methionyl and cysteinyl residues in the substrate  
 RT binding site of carboxypeptidase Y.";  
 RL Carlsberg Res. Commun. 49:639-645(1984).  
 RN [5]  
 RP ACTIVE SITE HIS-508.  
 RA MEDLINE=90315013; PubMed=2639680;  
 RX Bech L.M., Breddam K.;  
 RT "Inactivation of carboxypeptidase Y by mutational removal of the  
 RT putative essential histidyl residue.";  
 RL Carlsberg Res. Commun. 54:165-171(1989).  
 RN [6]  
 RP MUTAGENESIS.  
 RX MEDLINE=94114535; PubMed=7904479;  
 RA Mortensen U.H., Remington S.J., Breddam K.;  
 RT "Site-directed mutagenesis on (serine) carboxypeptidase Y. A hydrogen  
 RT bond network stabilizes the transition state by interaction with the  
 RT C-terminal carboxylate group of the substrate.";  
 RL Biochemistry 33:508-517(1994).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE=95244421; PubMed=7727362;  
 RA Endrizzi J.A., Breddam K., Remington S.J.;  
 RL "2.8-A structure of yeast serine carboxypeptidase.";  
 RT Biochemistry 33:11106-11120(1994).  
 CC -1- FUNCTION: INVOLVED IN DEGRADATION OF SMALL PEPTIDES. DIGESTS  
 CC PRESENTLY PEPTIDES CONTAINING AN ALIPHATIC OR HYDROPHOBIC  
 CC RESIDUE IN P1' POSITION, AS WELL AS METHIONINE, LEUCINE OR  
 CC PHENYLALANINE IN P1 POSITION OF ESTER SUBSTRATE.  
 CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a  
 CC broad specificity.  
 CC -1- ENZYME REGULATION: INHIBITED BY ZPCK.  
 CC -1- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.  
 CC -1- PTM: ENTERS THE ENDOPLASMIC RETICULUM AS AN INACTIVE ZYMOGEN AND  
 CC IS MODIFIED BY FOUR N-LINKED CORE OLIGOSACCHARIDES, GIVING RISE TO  
 CC A PRECURSOR KNOWN AS P1 (67 kDa). AS P1 TRANSITS THROUGH THE  
 CC GOLGI, EXTENSION OF ITS CORE OLIGOSACCHARIDES LEADS TO THE  
 CC GOLGI-MODIFIED P2 PRECURSOR (69 kDa). P2 IS SORTED AWAY FROM  
 CC SECRETORY PROTEINS AT OR BEYOND A LATE GOLGI COMPARTMENT AND IS  
 CC SUBSEQUENTLY DELIVERED TO THE VACUOLE VIA A PREVACUOLAR  
 CC ENDSOME-LIKE COMPARTMENT. UPON ARRIVAL IN THE VACUOLE, THE  
 CC N-TERMINAL PROSEGMENT OF P2 IS CLEAVED TO YIELD THE ENZYMATICALLY  
 CC ACTIVE MATURE VACUOLAR FORM OF CPY (61 kDa).





Db 326 ERCLGLIESCDYSDQ-----SVMSCVFATYYCNNAQLAPYOR-TGRNVYDIRKDCGGNLC 379

QY 348 -TPPSYNNKFLAKDSVDAIGNYNTOS-NDVYVAFQOQDGFVWPNFIEDLELIAL 405

Db 380 YTLQD-IDYLNQDYVKGAEVADHYHSCNFIDNRNLFAGDWKPK-YHTAVTDLLNQ 437

QY 406 PYRSLIYGDADYICNWFQGVASLAANTSSQAQFRS---AGYTPLKVGVEYGETREYG 462

Db 438 DLPILYVAGDKDFICNMLGNKAWTDVLPWKYDEFAFASQVRNWTASITIDEVA-GEVYSYK 496

QY 463 NFSFTRVYAGHEVPPYQPIASLQLENRIFG 494

Db 497 HFTYLRVFGGHMVPDPVPENALSMVNEWIHG 528

## RESULT 7

CBPX ARATH STANDARD; PRT; 516 AA.

AC P32826; Q42107; Q9CAE5;

DT 01-OCT-1993 (Rel. 27, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Serine carboxypeptidase precursor (EC 3.4.16.-).

GN AT3G10410 Or F3JM14.32.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Bradley D.;

RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=21016720; PubMed=11130713;

RA Salanoubat M., Lemcke K., Rieger M., Ansorge W., Unseld M.,

RA Farmanb B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,

RA Delsen M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,

RA De Simone V., Choise N., Artiguenave F., Robert C., Brottier P.,

RA Winkler P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,

RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,

RA Wurba E., Dzionek H., Erle H., Jordan N., Bangert S.,

RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,

RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,

RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,

RA Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,

RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,

RA Cooke R., Laudie M., Berger-Lilauro C., Purnelle B., Masuy D.,

RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,

RA Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,

RA Manhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,

RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,

RA Rooney T., Rizzo M., Waits A., Utterback T., Fujii C.Y., Shea T.P.,

RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,

RA Pai G., Millscher J., Sellers P., Gill J.E., Feldblum T.V.,

RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,

RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,

RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,

RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,

RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,

RA Watanabe A., Yamada M., Yasuda M., Tabata S.;

RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana".

RL Nature 408:820-822(2000).

RP [3]

RN SEQUENCE OF 252-372 AND 455-516 FROM N.A.

RC STRAIN=cv. Columbia; TISSUE=Seedling;

RA Hofte H.;

RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.

CC -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a broad specificity.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; M81130; AAB04606.1; -

CC DR EMBL; AC011560; AAG51389.1; -

CC DR EMBL; Z25955; CAA81115.1; -

CC DR EMBL; Z26528; CAA81299.1; -

CC DR HSSP; P00729; IYSC.

CC DR MEROPS; S10.009; -

CC DR InterPro; IPR000379; Ser\_estrs\_site.

CC DR InterPro; IPR001563; Serine\_carbpept.

CC DR Pfam; PF00450; serine\_carbpept; 1.

CC DR PRINTS; PR00724; CRBOXYPTASEC.

CC DR ProDom; PD001189; Serine\_carbpept; 1.

CC DR PROSITE; PS00131; CARBOXYPEPT\_SER\_SER; 1.

CC DR PROSITE; PS00560; CARBOXYPEPT\_SER\_HIS; 1.

CC KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.

CC FT SIGNAL 1 22 POTENTIAL.

CC FT PROPEP 23 82 POTENTIAL.

CC FT CHAIN 83 516 SERINE CARBOXYPEPTIDASE.

CC FT ACT\_SITE 229 229 BY SIMILARITY.

CC FT ACT\_SITE 417 417 BY SIMILARITY.

CC FT ACT\_SITE 474 474 BY SIMILARITY.

CC FT BINDING 420 420 SUBSTRATE (BY SIMILARITY).

CC FT CARBOHYD 157 157 N-LINKED (GLCNAC...) (POTENTIAL).

CC FT DISULFID 139 379 BY SIMILARITY.

CC FT DISULFID 300 322 BY SIMILARITY.

CC FT DISULFID 307 315 BY SIMILARITY.

CC FT CONFLICT 515 516 QM -> ADVTTSFALHKNKQOIIKQCVSN (IN REF. 1).

CC SQ SEQUENCE 516 AA; 57301 MW; 177C778DF657A1C1 CRC64;

Query Match 16.4%; Score 474; DB 1; Length 516;

Best Local Similarity 30.1%; Pred. No. 6.le-26;

Matches 139; Conservative 71; Mismatches 172; Indels 80; Gaps 18;

QY 45 TTPGVKYSQY-VDTSPESHFFWFEARHNPETAPILWNGGPGSDSLGLPEELGP 102

Db 93 TVDDLGHHAGYKLPKSRGASMPFFESR-NKKDAPVVIWLGPGSCSELAFFYENGP 151

QY 103 CHVNSTFDDYINPHSNWEVSNLFLSQPLGVGESYSDTVGDSINPVTVGVVENSFAGVQG 162

Db 152 FKITSNNKSLAWNEYGDQVSNLLYVDQPVCTGFSYTTDKSDIHDETVG----- 200

QY 163 RYPTIDATLIDTNNLAAEAWEILQGLSLGSLDSRVOSKDFSLWTSYGGHYGPAFFN 222

Db 201 -----SNDLYDFLQAFFAHEPKL-----AKNDFYITGESYAGHYIPAFAS 240

QY 223 HEYEQNERIANGSVNGVQLNFNSLGLTINGLIDEAIOAPYPPEFVANNYTGIKAVNTYVN 282

Db 241 RVHKGNNK--AN--EGVHNLKGFAGNGLTDPALQYPAIPDYALE--MGLITQKE---- 289

QY 283 YMKFANOMPGNQDLTSTCKQTNRTA-LADYALCAEATNMCNDRNVGPGYAFAGRGVYDI 341

Db 290 HDRLEKIVPL-CELSIKLCTGDTTSCIASLYVNCNSLFSGVMSHAGGVNY-----YDI 341

QY 342 RHP-----YDDTTPSYNNKFLAKDSVMDAIGV-NINYTQSNNDVYVAFQOQDGFVWPN 394

Db 342 RKKCVGSLCYDF-----SNMEKFLNLQSVRKSLGVGDIDFVSCSTSVYQA-----MLVD 390

QY 395 FIEDLEEIILALPVRYS-----LIY-GDADYICNWFQGVASLAANTSSQAQFRSAGYTPL 448

Db 391 WMRNLE--VGIPITLLEDGISLLVYAGEYDLICNWLGNRSWVNAWMSGKTFNCAKEVPF 448

QY 449 KVGVEYGETREYGNFSTRVYAGHEVPPYQPIASLQLENR 490



```

FT PROPEP      22 73      BY SIMILARITY.
FT CHAIN       74 484
FT PROPEP     485 500      BY SIMILARITY.
FT ACT_SITE   216 216      BY SIMILARITY.
FT ACT_SITE   404 404      BY SIMILARITY.
FT ACT_SITE   461 461      BY SIMILARITY.
FT BINDING    407 407      SUBSTRATE (BY SIMILARITY).
FT CARBOHYD   144 144      N-LINKED (GLCNAC... (POTENTIAL).
SQ SEQUENCE   500 AA; 55446 MW; 554552780147DB8 CRC64;

Query Match      15.7%; Score 452; DB 1; Length 500;
Best Local Similarity 26.4%; Pred. No. 2.le-24;
Matches 128; Conservative 81; Mismatches 214; Indels 62; Gaps 14;

Qy 14 LPK--NPTGVKTLTANNVIRYKEPGAECETTPGVKSYSGY--VDTSPESHTFFWF 69
Dy 47 LPKEAGTGDVPSVAPGELLERRVTLPGLPQGVGLGHAGYRFLNTHDARMFYLE 106
Qy 70 EARHNPETAPITLWLNGSGSDSLIGLFEELGCHVNSTDDYINPHSNVSNLLFLSQ 129
Dy 107 ESRGKKED--PVVITLGTGGPCSSSELAVFYENGPTTISNNMSLAWKFGWDTISNLIIFVDQ 165
Qy 130 PLGVGFSYSDTVGDSINPVTGVVNSFSFAGVQGYRPTIDAFLDITNLAARAAWEILQGF 189
Dy 166 PTGTGFSYSSDDRDTRHDETV-----SNDIYSFLQVF 198
Qy 190 LSGPLSDSRVQSKDFSLWTSYSGHYGPAFFNFHYEQNERIANGSVGNLNFNSGLII 249
Dy 199 FKHPEF-----AKNDFITGESYAGHYTPAPASRVHOGNK--AN---EGIHINKLGEAIG 249
Qy 250 NGIIDEATAQPYPEFAVNNTYIGKAVNETVYNTMKTANOMPNSCODLISCTCKQTNRIAL 309
Dy 250 NGLTDPALQKAYTDYALD-----NMLIKSYDRI---NKFIPPCFAIKLCGGNGK--- 299
Qy 310 ADYALCAEATNMCNRDNEGPGYAFAGRVGYDIRHYDDPTPPSYN--KFLAKDSYMDAI 367
Dy 300 ---ASCAAAYVVCNSIFSSIMKLVTGKNYDVKCECKLCYDFSNLEKFFGDKAVKEAI 356
Qy 368 GV-NINTQSNNDVYPAQQTGDFWPNFIE-DLEEILALPVRSVLIYGDADYICNTPGG 425
Dy 357 GVGDLFEVSCSTTVYQAMLTD---WMKNLEVGIPALLEDGINVLIVAGEVDLNCNWLGN 412
Qy 426 QAVSLAANYCAAQFRAGYTPLVKNGVEYGETREYGNFSFTRVVEAGHEVPYTCPIASL 485
Dy 413 SRVHSHWSGQKDFVSSHESPFVVDGAEAGVLKSHGPLSLFKVNHAGHVPMQPKASL 472
Qy 486 QLFNR 490
Dy 473 EMLRR 477

RESULT 10
CBP3_HORVU STANDARD; PRT; 508 AA.
ID CBP3_HORVU
AC P21529;
DT 01-MAY-1991 (Rel. 18, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-MIII).
GN CBP3 OR CXP;3.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Himalaya; TISSUE=Aleutone;
RA Roher A., Lok F., Cameron-Millis V., von Wettstein D.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 81-491.
RC STRAIN=cv. Gula;

```

```

RX MEDLINE=90315015; PubMed=2639682;
RA Soerensen S.B., Svendsen I., Breddam K.;
RT "Primary structure of carboxypeptidase III from malted barley.";
RL Carlsberg Res. Commun. 54:193-202(1989).
CC -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -!- ENZYME REGULATION: INHIBITED BY MERCURIC IONS.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOSPERM.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED MAINLY IN THE ALEURONE AND, TO A
CC LESSER EXTENT IN THE EMBRYO, THROUGHOUT THE 5-DAYS GERMINATION
CC PERIOD EXCLUSIVELY, WITH A MAXIMAL LEVEL AT 3 DAYS. ALSO FOUND IN
CC THE ROOTS AND SHOOTS OF THE GROWING SEEDLING.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Y09604; CAA70817.1; -.
DR HSSP; P00729; LCPY.
DR MEROPS; S10.009; -.
DR InterPro; IPR000379; Ser_estrs_site.
DR InterPro; IPR001563; Serine_carbpept.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXPTASEC.
DR PRODOM; P001189; Serine_carbpept; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 19
FT PROPEP 20 80
FT CHAIN 81 491 SERINE CARBOXYPEPTIDASE III.
FT PROPEP 492 508
FT MOD_RES 81 81 BLOCKED.
FT ACT_SITE 223 223 BY SIMILARITY.
FT ACT_SITE 411 411 BY SIMILARITY.
FT ACT_SITE 468 468 BY SIMILARITY.
FT BINDING 414 414 SUBSTRATE.
FT CARBOHYD 151 151 N-LINKED (GLCNAC... (POTENTIAL).
FT VARIANT 265 265 Q -> V (POLYMORPHISM).
SQ SEQUENCE 508 AA; 56362 MW; 70C6751D78D40AB6 CRC64;

Query Match      15.0%; Score 434; DB 1; Length 508;
Best Local Similarity 26.5%; Pred. No. 3.9e-23;
Matches 132; Conservative 77; Mismatches 221; Indels 68; Gaps 15;

Qy 1 LPGSTPASVGRRLPKNETGVKTLTANNVIRYKEPG--AEGVCETTPGVKSYSGY--VD 57
Dy 47 LPKDSSSSSGRHG--ARVGEEDVAPGOLLERRVTLPGLPGEVAD----LGHAGYRFLP 101
Qy 58 TSPESHTFFWFPEARHNPETAPITLWLNGSGSDSLIGLFEELGCHVNSTEDDYINPHS 117
Dy 102 NTHDARMFYFFESGKKED--PVVITLGTGGPCSSSELAVFYENGFTTANNMSLVNKEG 160
Qy 118 WNEVSNLLFLSQPLGVGFSYSDTVGDSINPVTGVVNSFSFAGVQGYRPTIDAFLDITNLI 177
Dy 161 WDKISNIIFVDOPTGTGFGFSYSSDDRDTRHDETV----- 194
Qy 178 ARAAAWEILQGLSGLSDSRVQSKDFSLWTSYSGHYGPAFFNFHYEQNERIANGSVN 237
Dy 195 -SNDIYDFLQVFFKKHPEF-----IXNDFITGESYAGHYTPAPASRVHOGNK-----NE 244
Qy 238 GVQLNFNSGLIINGIIDEATAQPYPEFAVNNTYIGKAVNETVYNTMKTANOMPNSCODL 297
Dy 245 GTHINKGFAICNGLTDPALQYKAYTDYALENMLNLTQKADYERI-----NKFIPPCFA 297
Qy 298 ISTCKQTNR--TALADYALCAEATNMCNRVNEGPGYAFAGRVGYDIRHYDDPTPPSYN- 355

```

```

Db 298 IKLCGTNGKASCMAAYVVCNTIENSIMKLVGKNY-----YDVRKECEGKLCYDFSNL 350
QY 356 -KFLAKDSVMDAIGV-NTINYTOSNNVDVYAFQQTGDFVWPNFIE-DLEEIIIALPVKVSILI 412
Db 351 EKFGDKAVRQIGVGDIEFYVSCSTSVQAMLT-----WMRNLEVGIPALLEDGINVLIY 406
QY 413 YGDADYICNWRGGQAVSLAANYSQAAQFRSAGYPLKYNVNGVEYGTREYGNFSTRVYDA 472
Db 407 AGEVDLICNWLGNRWHMSMWSGOKFAKTAESSFLYDDAQAGVLKSHGALSFLKVHNA 466
QY 473 GHEVPYQPIASLQLFNR 490
Db 467 GHMVPMDQPKAALMLRR 484

RESULT 11
CBP3_WHEAT
ID CBP3_WHEAT STANDARD; PRT; 500 AA.
AC P11515;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y.";
RL J. Biol. Chem. 262:13726-13735(1987).
CC -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -!- SUBUNIT: Monomer (Probable).
CC -!- INDUCTION: By gibberellic acid (GA).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02817; AAA34273.1; -.
DR FIR; A29412; A29412.
DR HSP; P00729; ICPY.
DR MEROPS; S10.009; -.
DR InterPro; IPR000379; Ser_estrs_site.
DR InterPro; IPR001563; Serine_carpept.
DR Pfam; PF00450; serine_carpept; 1.
DR PRINTS; PR00724; CREOXIPASEC.
DR ProDom; PD001189; Serine_carpept; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 21
FT PROPEP 22 73
FT CHAIN 74 484
FT PROPEP 485 500
FT ACT_SITE 216 216
FT ACT_SITE 404 404
FT ACT_SITE 461 461
FT BINDING 407 407
FT CARBOHYD 144 144
SQ SEQUENCE 500 AA; 55334 MW; B2ACE10EF8484CDA CRC64;

```

```

Query Match 14.3%; Score 413.5; DB 1; Length 500;
Best Local Similarity 25.7%; Pred. No. 1e-21;
Matches 122; Conservative 75; Mismatches 201; Indels 77; Gaps 15;

QY 38 GAEGVCE-----TTPGVKS-----YSGY--VDYSPESHTEFWFEARHNPEAPI 80
Db 58 GAEDVAPQQLERRVTLPLPGVGDGLHHAGYKLPNTHDARMEYFFESRGKED-PV 116
QY 81 TILWNGGSDSLIGLFEELGCHVNSTFDYINPNSWNEVSNLFLSQPLGVGFSYSDT 140
Db 117 VILWLTGGCCSELAVFYENGFTTANNMSLVNKGWDKISNLIIVDPATGTGFSYSD 176
QY 141 VDGSTNPVTGVVENSFAGVGQRYPTIDATLDTTNLAAEAAWEILQGLSLGSLDSRV 200
Db 177 DRDTRHDEAGV-----SNDLYDQLQVFKKHPF----- 205
QY 201 QSKDFSLWTSYGHGYPAPFNHFYEONERIANGSVNGVQVNFNSLGIINGLIDRAIQAP 260
Db 206 VKNDEFITGESYAGHYIEAFASRVHGNKK-----NEGTHINKLGAIGNGLDPAIQK 260
QY 261 YYPEFAVNTYGIKAVNETVYNYMKFANOMPNGCODLSTCKOTNR-TALADYALCAEAT 319
Db 261 AYTDYALD-----MNLIQADYDRI---NKFIPPCEFAIKLCGTGDKASCMAAYVNCNIF 313
QY 320 NMCNDNVGPPYAFAGRGVYDIRHPYDDPTPESYN--KFLAKDSVMDAIGV-NINYTQS 376
Db 314 NSIMKLVGKNY-----YDVRKECEGKLCYDFSNLEKFFGDKAVRQIGVGDIEFYSC 366
QY 377 NNDVYVAFQQTGDFVWPNFIE-DLEEIIALPVRSLLYGDADYICNWRGGQAVSLAANY 435
Db 367 STSVYQAMLT-----WMRNLEVGIPALLEDGINVLIYAGEVDLICNWLGNRWHMSMWS 422
QY 436 QAAQFRSAGYPLKYNVNGVEYGTREYGNFSTRVYAGHEVPYQPIASLQLFNR 490
Db 423 GQKDFAKTAESSFLYDDAQAGVLKSHGALSFLKVHNGHMPMDQPKAALMLRR 477

RESULT 12
CPVL_HUMAN
ID CPVL_HUMAN STANDARD; PRT; 476 AA.
AC Q9H3G5; Q9NB17; Q96AR7; Q9HB41;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable serine carboxypeptidase CPVL precursor (EC 3.4.16.-)
DE (Carboxypeptidase, vitellogenic-like) (Vitellogenic carboxypeptidase-
DE like protein) (VCP-like protein).
GN CPVL OR VLP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21295045; PubMed=11401439;
RA Mahoney J.A., Ntcolosi B., DaSilva R.P., Gordon S., McKnight A.J.;
RT "Cloning and characterization of CPVL, a novel serine
RT carboxypeptidase, from human macrophages.";
RL Genomics 72:243-251(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Cho J.-J., Baik H.-H.;
RT "Cloning of VCP-like protein expressed in human heart and placenta.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Tissue-Placenta;
RC Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,
RA Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,
RA Nagahari K., Sugano S., Isoqai T.;
RA "HRI human cDNA sequencing project.";
RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]

```

RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=22389257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.L., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McWan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Tuchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.J., Skalska U., Smailus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: May be involved in the digestion of phagocytosed  
 CC particles in the lysosome, participation in an inflammatory  
 CC protease cascade, and trimming of peptides for antigen  
 CC presentation.  
 CC -1- TISSUE SPECIFICITY: Expressed in macrophages but not in other  
 CC leukocytes. Abundantly expressed in heart and kidney. Also  
 CC expressed in spleen, leukocytes, and placenta.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF106704; AAC37991.2; -  
 CC EMBL: AF282617; AAG1348.1; -  
 CC EMBL: AK075433; BAC11618.1; -  
 CC EMBL: BC016838; AAH16838.1; -  
 CC HSSP: P10619; IIVY.  
 CC MEROPS: S10.003; -.  
 CC Genew: HGNC:14399; CPVL.  
 CC InterPro: IPR000379; Ser\_estr\_site.  
 CC InterPro: IPR001563; Serine\_carbpept.  
 CC Pfam: PF00450; serine\_carbpept; 1.  
 CC PRINTS: PR00724; CRBOXPASEC.  
 CC ProDom: PD001169; Serine\_carbpept; 1.  
 CC PROSITE: PS00131; CARBOXYPEPT\_SER\_SER; 1.  
 CC PROSITE: PS00560; CARBOXYPEPT\_SER\_HIS; FALSE\_NEG.  
 CC Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.  
 KW SIGNAL  
 FT SIGNAL 1 22  
 FT PROPEP 23 ? POTENTIAL.  
 FT CHAIN ? 476 PROBABLE SERINE CARBOXYPEPTIDASE CPVL.  
 FT ACT\_SITE 204 204 BY SIMILARITY.  
 FT ACT\_SITE 388 388 BY SIMILARITY.  
 FT ACT\_SITE 448 448 BY SIMILARITY.  
 FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 346 346 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 25 25 R -> H (IN REF. 2).  
 FT CONFLICT 284 284 L -> F (IN REF. 3 AND 4).  
 FT CONFLICT 287 287 F -> L (IN REF. 3).  
 FT CONFLICT 398 398 H -> R (IN REF. 3 AND 4).  
 FT CONFLICT 422 422 F -> L (IN REF. 2).  
 FT CONFLICT 435 435 A -> V (IN REF. 3 AND 4).  
 FT CONFLICT 438 438 F -> S (IN REF. 2).  
 CC SEQUENCE 476 AA; 54110 MW; 2D966683A4F3FD01 CRC64;

Query Match 12.4%; Score 359; DB 1; Length 476;  
 Best Local Similarity 26.7%; Pred. No. 6.5e-18;  
 Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;  
 QY 47 PG--VKSYSGY--VDTSPESHTEFFEFARHNPETAPITLWNGGPGSDSLIGLFEELGP 102  
 DB 67 PGLNMSYAGFLIVNKYNSNLEFFEPQAQIQEDAPVVLWLOGGPGSSMEGLFVEHGP 126  
 QY 103 CHVNSTEDDDYINPHSNWNEVSNLLFLSPLQGVGSYSYSDTVGDSINPVTVGVVENSFAGVQG 162  
 DB 127 YVTSNTMLRDRDPFWTILSMLYIDNPVGTGFSFTDDTHG----- 167  
 QY 163 RYFTIDATLDTNLAEEAAWEILQGLSLPILDSRVOSKDFSLWTSYGGHYGPAFFN 222  
 DB 168 -----YAVNEDDVARLYSALIQFF-----QTFPEYKNDFFVIGESTAGKYVPAIAH 215  
 QY 223 HEYEQNERIANGSVNGVQLNFNSIGITINGIIDEAICAPYYPPEFVANNIYIGIKAVNETVYN 282  
 DB 216 LIHSLN-----PVREVKINLNGTALGDYSDPEIIGYAEF-----LYQIGLLDEKQK 265  
 QY 283 YMFANOMPNGCQQLISTCKQTN---RTALADYALCAEATN---MCRDNVEG--PYAFAG 335  
 DB 266 Y--FQKQ---CHECHIRKQNWLEAFELDKLLDGLTSDPSYFQNVTCGSNYNFL- 318  
 QY 336 RGVYDIRHPYDDPTPPSYNKLAKDSVMDAIGNVNTQSNNDVY--YAFQOTGDFVWP 393  
 DB 319 -----RCTEPEDQLYYVKFLSLPEVROAIHVG--NOTFNDGTIVVEKYRLREDIVQSVKP 369  
 QY 394 NFIEDLEILALPVRVSLIYGDADYI-----CNWFGSQAVSLAANSQAQ-- 439  
 DB 370 WLFEIMNN-----YKVLIIYNGQLDIIVAAALTEHSLMGMDWKGSQ-----EYKKAEEKV 418  
 QY 440 ---FRS-----AGYTPKLVNGVEYGETREYGNFSTRVYEGAEHVEVPYOPATSLQFNRI 492  
 DB 419 WKIPKSDSEVAGY-----IRQAGDFHOVIIRGGSHILPYQPLRAFDMINRFI 466  
 QY 493 F--GWD 496  
 DB 467 YGKGWD 472  
 RESULT 13  
 CP22\_HORVU  
 ID CP22\_HORVU STANDARD; PRT; 436 AA.  
 AC P55748;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Serine carboxypeptidase II-2 precursor (EC 3.4.16.6) (CP-MII.2)  
 DE (Fragment).  
 DE CXP;2-2.  
 GN Hordeum vulgare (Barley).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
 OC Triticeae; Hordeum.  
 OC NCBI\_TaxID=4513;  
 OX [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RN STRAIN=cv. Alexis; TISSUE=Grain;  
 RX MEDLINE=94336715; PubMed=7520177;  
 RA Dal Degun F., Rocher A., Cameron-Mills V., von Wettstein D.;  
 RT "The expression of serine carboxypeptidases during maturation and  
 RT germination of the barley grain.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:8209-8213(1994).  
 CC -1- CATALYTIC ACTIVITY: Preferential release of a C-terminal arginine  
 CC or lysine residue.  
 CC -1- SUBUNIT: CARBOXYPEPTIDASE II IS A DIMER, WHERE EACH MONOMER IS  
 CC COMPOSED OF TWO CHAINS LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE GERMINATING EMBRYO. LOW  
 CC LEVELS IN THE DEVELOPING ALEURONE AND EMBRYO. ALSO FOUND IN THE  
 CC ROOTS AND SHOOTS OF THE GROWING SEEDLING.  
 CC -1- PTM: THE LINKER PEPTIDE IS ENDOPROTEOLYTICALLY EXCISED DURING

CC ENZYME MATURATION (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X78878; CAB59202.1; -;  
 CC HSP: P08819; IWHI.  
 CC InterPro: IPR000379; Ser\_estrs\_site.  
 CC Pfam: PF00450; serine\_carbpept.  
 CC PRINTS: PR00724; CRBOXYPTASEC.  
 CC ProDom: PD001189; Serine\_carbpept; 1.  
 CC PROSITE: PS00131; CARBOXYPEPT\_SER\_SER; 1.  
 CC PROSITE: PS00560; CARBOXYPEPT\_SER\_HIS; 1.  
 CC KW Hydroxylase; Carboxypeptidase; Glycoprotein; Zymogen; Multigene family.  
 CC NON\_TER 1  
 CC CHAIN <1 256  
 CC PROPEP 257 270 SERINE CARBOXYPEPTIDASE II-2, CHAIN A.  
 CC CHAIN 271 436 LINKER PEPTIDE (BY SIMILARITY).  
 CC ACT\_SITE 149 149 SERINE CARBOXYPEPTIDASE II-2, CHAIN B.  
 CC ACT\_SITE 350 350 BY SIMILARITY.  
 CC ACT\_SITE 403 403 BY SIMILARITY.  
 CC DISULFID 56 313 INTERCHAIN (BY SIMILARITY).  
 CC DISULFID 217 229 BY SIMILARITY.  
 CC DISULFID 253 281 BY SIMILARITY.  
 CC CARBOHYD 107 107 N-LINKED (GLCNAC... ) (POTENTIAL).  
 CC SEQUENCE 436 AA; 48952 MW; E0F82D97E0C34DC9 CRC64;  
 CC -----  
 CC Query Match 12.2%; Score 353.5; DB 1; Length 436;  
 CC Best Local Similarity 25.7%; Pred. No. 1.4e-17;  
 CC Matches 123; Conservative 76; Mismatches 173; Indels 107; Gaps 23;  
 CC -----  
 CC 52 YSGYVDTSP--SHTFFWFEARHNPTAPITLWNGPGSDSL-IGLFEELGPGCHVNST 108  
 CC 16 YAGYVTSERDGAALFYWFEEAHDFAKPLLLWNGPGGCSSTAFGVEGVGFHVNAD 75  
 CC 109 FDD-YINPHSWNEVSNLLFSLQPLGVGFSDTVDGSIINPTGVVNSSFAGYQGRYPTI 167  
 CC 76 GKGVHNPYSWNVANILFDSPGVGYSYNT-----SADILSN-----GDRTA 121  
 CC 168 DATLIDTTLAABAWEILQGLFSLGFLSDRSVQSKDFSLWTSYGYGHYGPAPFNHFEQ 227  
 CC 122 KDSLVELTK-----N-LERF-----PQYKEREFLYTGSGYAGHYVPQLAQAIKRH 165  
 CC 228 NERIANGSVNGVNLNSLGIINGIIDEAIQAPYPEFAVNNYGIKAVNETVYNNMKA 287  
 CC 166 HEATGDKSI-----NLKGYMVGNAITDDF--HDHYGIFQYMTTGL--ISDQYKLLNF 216  
 CC 288 NOMPNGC--QDLISTCKQNRATADYALCAEATNMCRCNVEGP--YYAFAG----- 335  
 CC 217 -----CDPESFHTSPQDK--ILDIA--STEAGNIDSYSIPTTCHSFSASSRNKVKR 267  
 CC 336 -RGVYDIRHPYDDTPPS---YNNKFLAKDS--VMDAIG-----VNINYTGSSND 379  
 CC 268 LRSVKGMEQYDPCTEKHSIVFNLEHVQKALHVNVPVIGKSKWETCSEVININWKDCRS 327  
 CC 380 VYFAFOQTGDFVWPNEIDLEELLALPVRSYLIIGDADYICNWFPGQAVSLAANYSOAAQ 439  
 CC 328 VLIHY-----HEITQGLRIWMPSGDTDAV-----IPVSTSTRYSDA 364  
 CC 440 FRSAGYTPLVK-----NGVEYGETREYCNFSFTRVYEGAGHEVPYQPIASLQLFNRTIFG 494  
 CC 365 LXLPTVTPHWAVDDGEGVGWVGQYKGLNFVIVRGAGHEVPLHRPKQALTLIKSLIAG 423

RESULT 14  
 YSS2\_CAEEL

ID YSS2\_CAEEL STANDARD; PRT; 470 AA.  
 AC Q09991;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative serine carboxypeptidase K10B2.2 precursor (EC 3.4.16.-).  
 GN K10B2.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Miller N.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U28730; AAA68259.1; -;  
 CC PIR: T16606; T16606.  
 CC HSP: P10619; IIVY.  
 CC MEROPS: S10.002; -;  
 CC WormPep: K10B2.2; CE02009.  
 CC InterPro: IPR000379; Ser\_estrs\_site.  
 CC InterPro: IPR001563; Serine\_carbpept.  
 CC Pfam: PF00450; serine\_carbpept; 1.  
 CC PRINTS: PR00724; CRBOXYPTASEC.  
 CC ProDom: PD001189; Serine\_carbpept; 2.  
 CC PROSITE: PS00131; CARBOXYPEPT\_SER\_SER; 1.  
 CC PROSITE: PS00560; CARBOXYPEPT\_SER\_HIS; 1.  
 CC KW Hypothetical protein; Hydrolase; Carboxypeptidase; Glycoprotein;  
 CC Signal.  
 CC SIGNAL 1 19 POTENTIAL.  
 CC CHAIN 20 470 PUTATIVE SERINE CARBOXYPEPTIDASE K10B2.2.  
 CC ACT\_SITE 169 169 BY SIMILARITY.  
 CC ACT\_SITE 380 380 BY SIMILARITY.  
 CC ACT\_SITE 441 441 BY SIMILARITY.  
 CC CARBOHYD 132 132 N-LINKED (GLCNAC... ) (POTENTIAL).  
 CC CARBOHYD 316 316 N-LINKED (GLCNAC... ) (POTENTIAL).  
 CC CARBOHYD 396 396 N-LINKED (GLCNAC... ) (POTENTIAL).  
 CC SEQUENCE 470 AA; 53158 MW; CCC2DACB75EF30FC CRC64;  
 CC -----  
 CC Query Match 12.2%; Score 352.5; DB 1; Length 470;  
 CC Best Local Similarity 26.4%; Pred. No. 1.8e-17;  
 CC Matches 131; Conservative 60; Mismatches 179; Indels 127; Gaps 20;  
 CC -----  
 CC 46 TPGVKSYSGVDTSPESHTEFFWFEARHNPTAPITLWNGPGSDSLIGLFEELGPGCHV 105  
 CC 35 TPDFHYSGYIRAWTDKYLHWLTESRSRAPTDQPLVLWNGPGGCSLDGLIELGPGFHV 94  
 CC 106 NSTFDD-YINPHSWNEVSNLLFSLQPLGVGFSDTVDGSIINPTGVVNSSFAGYQGRY 164  
 CC 95 KDFGNSIYYNEYAWNKFAVNFLESLPAGVGYSTINFLTVSDDEVSLN----- 144  
 CC 165 PTIDATLIDPTNLAAEAAWEILQGLFSLGFLSDRSVQSKDFSLWTSYGYGHYGPAPFNH 224  
 CC 145 -----YMALDLFLSKPEYKGR---DFWITGESYAGVIPTL--- 178  
 CC 225 YEONERIANGSVNGVNLNSLGIINGIIDEAIQAPYPEFAVNNYGIKAVNETV 280  
 CC 179 ---AVRILNDKKNP--NFKGVAIGNALNFPNNYNTMVPFY-----YYHALVRDDL 225  
 CC 281 YNYMKEANQMPNGCQDLISTCKQNRATADYALCAEATNMCRCNVEGYPYAFAGRGVYD 340  
 CC 226 YN-----DIARNCCNNIGTCD-----IYKFFDPN--CRDKV---INALDGTNELN 267



*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 22:37:42 ; Search time 78 Seconds  
(without alignments)  
1776.593 Million cell updates/sec

Title: US-09-712-338-2\_COPY\_19\_555  
Perfect score: 2887  
Sequence: 1 LPOSTPASVGRQLPKNP7G.....HTQSSVPLPTATSMSSVGMA 537

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2884	99.9	555	3 Q96VZ9	Q96vz9 aspergillus
2	746.5	25.9	460	3 Q12569	Q12569 absidia zyc
3	601	20.8	552	3 Q96VC4	Q96vc4 emericeella
4	596.5	20.7	541	3 Q94152	Q94152 pichia angu
5	595.5	20.6	537	3 Q14414	Q14414 pichia angu
6	479	16.6	507	10 Q8VWQ0	Q8vwq0 gossypium h
7	473	16.4	429	10 Q8GV71	Q8gv71 oryza sativ
8	466.5	16.2	505	10 Q9FFB0	Q9ffb0 arabidopsis
9	461	16.0	501	10 Q9XH61	Q9xh61 matricaria
10	460	15.9	508	10 Q8L6A7	Q8l6a7 theobroma c
11	439	15.2	510	10 Q9LXH4	Q9lxh4 arabidopsis
12	438	15.2	510	10 Q93ZC3	Q93zc3 arabidopsis
13	413.5	14.3	2105	5 Q17679	Q17679 caenorhabdi
14	406.5	14.1	499	10 Q9FMX9	Q9fmx9 arabidopsis
15	385	13.3	490	10 Q9FPY7	Q9fpy7 oryza sativ
16	383	13.3	482	10 Q9SV04	Q9sv04 arabidopsis

17	376.5	13.0	524	10 Q8GTK2	Q8gtk2 oryza sativ
18	363	12.6	480	10 Q9LEY1	Q9ley1 arabidopsis
19	360	12.5	437	10 Q9FWG1	Q9fwg1 oryza sativ
20	360	12.5	482	10 Q9FRJ0	Q9frj0 oryza sativ
21	359	12.4	487	10 Q9SV03	Q9sv03 arabidopsis
22	357.5	12.4	479	10 Q949Q7	Q949q7 arabidopsis
23	355.5	12.3	360	10 Q9M450	Q9m450 cicer ariet
24	353	12.2	470	10 Q22732	Q22732 arabidopsis
25	349	12.1	474	11 Q9D2D1	Q9d2d1 mus musculu
26	348	12.1	452	10 Q9ZQO0	Q9zqo0 arabidopsis
27	344	11.9	465	10 Q04084	Q04084 arabidopsis
28	344	11.9	494	10 Q9FEU4	Q9feu4 pisum sativ
29	344	11.9	501	10 Q9SV02	Q9sv02 arabidopsis
30	342.5	11.9	471	10 Q8L9Y0	Q8l9y0 arabidopsis
31	342.5	11.9	473	10 Q9SCA9	Q9sc9a arabidopsis
32	342.5	11.9	669	10 Q8L9Y6	Q8l9y6 oryza sativ
33	340	11.8	512	5 Q76725	Q76725 caenorhabdi
34	337.5	11.7	425	10 Q65568	Q65568 arabidopsis
35	333.5	11.6	510	3 Q60123	Q60123 schizosacch
36	333	11.5	487	10 Q9ZUG3	Q9zug3 arabidopsis
37	332	11.5	465	10 Q9M099	Q9m099 arabidopsis
38	331	11.5	497	10 Q8L7B2	Q8l7b2 arabidopsis
39	330.5	11.4	472	10 Q9LSM9	Q9lsm9 arabidopsis
40	330.5	11.4	474	10 Q82228	Q82228 arabidopsis
41	330	11.4	445	5 Q814E3	Q814e3 caenorhabdi
42	330	11.4	1203	5 Q45916	Q45916 caenorhabdi
43	328.5	11.4	548	4 Q9BR08	Q9br08 homo sapien
44	328	11.4	504	10 Q9LSV8	Q9lsv8 arabidopsis
45	327.5	11.3	2338	5 Q94269	Q94269 caenorhabdi

#### ALIGNMENTS

#### RESULT 1

Q96VZ9 PRELIMINARY: PRT: 555 AA.  
AC Q96VZ9;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Carboxypeptidase S1 precursor (EC 3.4.16.6).  
GN CPI.  
OS Aspergillus oryzae.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5062;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TK3;  
RA van den Broek P.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF394242; AAK77166.1; -  
DR InterPro; IPR001563; Serine\_carbpept; 1.  
DR Pfam; PF00450; serine\_carbpept; 1.  
DR PRINTS; PR00724; CRBOXYPTASEC.  
DR ProDom; PD001189; Serine\_carbpept; 1.  
DR PROSITE; PS00131; CARBOXYPEPT\_SER\_SER; 1.  
KW Carboxypeptidase; Hydrolase; Signal.  
FT SIGNAL 1  
FT POTENTIAL 18  
SQ SEQUENCE 555 AA; 61168 MW; 456863B0CEB55222 CRC64;

Query Match 99.9%; Score 2884; DB 3; Length 555;  
Best Local Similarity 99.8%; Pred. No. 1.1e-195;  
Matches 536; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 LPGSTPASVGRQLPKNP7GKTLTANNVITRYKEPGAEGVCETPGVKSYSGYVDISP 60  
Db 19 LPGSTPASVGRQLPKNP7GKTLTANNVITRYKEPGAEGVCETPGVKSYSGYVDISP 78  
Qy 61 ESHTTFWFFEARHNPETAPITLWNLGGPGSDSLIGLFEELGPGCHVNSTFDDYINPHSWNE 120  
Db 79 ESHTTFWFFEARHNPETAPITLWNLGGPGSDSLIGLFEELGPGCHVNSTFDDYINPHSWNE 138



171	NGBPCCSSLTGMLELFGSASTGPDPLKPTNNPYSWNSNATVIFDQPVNVGFSYS-----224	Db
145	INPVTGVVENSFAGVQGRYPTIDATLLDTTNLAEEAAWEILLQGLSGLPDSLSRVQSKD204	QY
225	-----SKSYSNIVAAGKDVIYAFLELVFQQPPHL-----LKND256	Db
205	FSLWTESYGGHYGPAFFNHFEQNERIANGSVNGVOLFNPSLGIINGLIIDEALQAPYYPE264	QY
257	PHIAGESYGGHYIVPFASEILTHADR-----SFNLTSVLINGLGTDLNLPQPFYFR307	Db
265	FAVNNTYGIKAVNETVYNYMKFANQMPNGCQDLSTCKQTNRKTALADYALCAEATNMCR323	QY
308	MACSTDGGGYETDLESECEGM--LETLP-CLSLTESYSSQ-----SVFSCVPASLIYN359	Db
324	DNVEGPPYAFAGRGYDILRHP-----YDDTPPSPYNNKFLAKDSVMDAIGNI-NYTOS376	QY
360	NAQLGPFQK-TGRNVDYVRKMCCEGLCYKD--MEYIDQYLNQDVKKEKIGAEVETYESC415	Db
377	NDVYVYAFQNGDFWVP---NFIEDLEILALPVRVSLIY-CDADYICNWFQGVQAVSLAA432	QY
416	NFDYNRNRFLEAGDWMKPYHKVINLEQ--GLPQV---LTYAGDKDFICNWLNGQAWSNEL470	Db
433	NYSCAAQFRSAG-YTPLVKNGVEYGETREYGNFSTRVVYAGHEVPYQPIASQLQFNRT491	QY
471	PWSGHEEFESAEILNYLLTKDGTKYGEVKNAGKFTFARMPDGGHMVPYDQPESSLAMVNRW530	Db
492	IFGWDIAGBQKK503	QY
531	IAG-DYSLGTRKK541	Db

```

RESULT 5
O14414 PRELIMINARY; PRT; 537 AA.
ID O14414
AC O14414;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 23, Last annotation update)
DE Carboxypeptidase Y.
DE GN CPY.
OS Pichia angusta (Yeast) (Hansenula polymorpha).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_taxid=4905;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN-DL-1;
RA Bae J.-H., Kim H.-Y., Sohn J.-H., Choi E.-S., Rhee S.-K.;
RT "Molecular cloning and characterization of a gene coding for
RL carboxypeptidase Y from yeast Hansenula polymorpha DL-1.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DDBB databases.
DR EMBL: U671174; AAB68520.2; -.
DR HSSP: P00729; ICPE.
DR MEROPS: S10.001; -.
DR InterPro: IPR001563; Serine_carpept.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00450; serine_carpept; 1.
DR PRINTS: PR00724; CRBOXYPTASEC.
DR PRODOM: PD001189; Serine_carpept; 1.
DR PROSITE: PS00131; CARBOXYPEPT_SER_SER; 1.
SQ SEQUENCE 537 AA; 60270 MW; FCQ1E57D7582C941 CRC64;

```

	Query Match	20.6%	Score 595.5	DB 3	Length 537
	Best Local Similarity	32.3%	Pred. No. 8.1e-34		
	Matches 165	Conservative 65	Mismatches 196	Indels 85	Gaps 20
QY	13	OLKNPTGVKLTLTANNVTIRYK	-----EPGAEGVCEITPGVKSYGYVDTSPE-SHFF	65	
DB	92	KITKKPADEFWNVUNEFKNYKLRVKKYKTDPCALGLDHI	----ROYSGVLDEDEKHEF	147	
QY	66	FWFTEARHNPTAPITLWINGPGSDSLGLGLEELGPGCHVNSFTDDY	INPHSWNEVSNLL	125	
DB	148	YWMFESRNDPNVDPIYLWINGPGGSSITGMFLFELGASIGPDLKPINPY	SNWNSNATVI	207	



DB 166 K-----NSEGHINLKGAINGLTDPAIQKAYTDYSLD-----MGLTKSOFNFI---N 213  
 QY 289 QMPNGCQDLISTCKQTNR-TALADYALCAEATNMCRDNVEGPPYAFAGGVYDIRHP--- 344  
 DB 214 KIVPTCELAIAKLCGTSGTISGIGAYVVC-----NLIFFSIE---TIIGKKNYDIRKPCVG 266  
 QY 345 ---YDDPTPPSYNKFELAKDSVMDAIGV-NINYTQSNNDVYAFQOTGDFWPNFIE-DL 399  
 DB 267 SLCYD-----LSNMEKFIQLKSVRESLGVGDIOFVSCSTVTOAML-----LDMRNLVYGI 318  
 QY 400 BEILALPVRLSYLIYGDADYICNWFEGGQAVSLAANYSOAQRSGYTPKLVNGVYGETR 459  
 DB 319 PELLENDIKVLIYAGEYDLICNWLGNRSWNSMEMSGKEAFVSSREBFTVDGKAGILK 378  
 QY 460 EYGNFSTRVYAEAGHEVPYIOPIASLOLFNRTIFG 494  
 DB 379 SYGLPSFLKHVDAGHMVPMQPKVALEMLMRWTS 413

## RESULT 8

Q9FFB0 PRELIMINARY; PRT; 505 AA.  
 AC Q9FFB0;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Serine carboxypeptidase.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 ON NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=9741969; PubMed=9330910;  
 RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
 RA Miyajima N., Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
 RT features of the 1.6 Mb regions covered by twenty physically assigned  
 RT pl clones";  
 RL DNA Res. 4:215-230(1997).  
 DR EMBL; AB005243; BAB10619.1;  
 DR HSSP; P10619; IIVY.  
 DR MEROPS; S10.009;  
 DR InterPro; IPR001563; Serine\_carbpept.  
 DR InterPro; IPR000379; Ser\_estrs\_site.  
 DR Pfam; PF00450; serine\_carbpept; 1.  
 DR PRINTS; PR00724; CRBOXYPTASEC.  
 DR ProDom; PD001189; Serine\_carbpept; 1.  
 DR PROSITE; PS00560; CARBOXYPEPT\_SER\_HIS; 1.  
 DR PROSITE; PS00131; CARBOXYPEPT\_SER\_SER; 1.  
 KW Carboxypeptidase.  
 SQ SEQUENCE 505 AA; 56543 MW; 6E74CD351090B099 CRC64;

Query Match 16.2%; Score 466.5; DB 10; Length 505;  
 Best Local Similarity 28.3%; Pred. No. 9.9e-25;  
 Matches 132; Conservative 75; Mismatches 178; Indels 81; Gaps 15;  
 QY 47 PGVKS---YSGY--VDTSPESHITFFWFFEARHNPETAPITLWNGPGSDSLIGLFEELG 101  
 DB 91 PSVOEFCHVAGYSYSLPHSKAKMFYFFESR-NKTTDPVVVILTGGPGCSSVAMFYENG 149  
 QY 102 PCHVNSTFDYINPHSWNEVSNLLFLSQPLGVGFSDYTDVDSINPVTGVWNSFAGVQ 161  
 DB 150 PFKISKDLSLYWDFDQWKVSNIIYVDPVGTGFSYTSDESIRNDEGV----- 199  
 QY 162 GRYPTIDATLIDTTLAAEAWEILQGLSLSDRSVQSKDFSLWTSYGGHYGPAFF 221  
 DB 200 -----SNDYDLQFAFFKEHP-----KFVKNDFFITGESYAGHYIPALA 238  
 QY 222 NHYEQNERIANGSVNGVQLNFNSLGLINGIIDEAIOAPYYPEFVNNYVIGIKAVNEIVY 281

DB 239 SRVHSGNKK-----KEGIPINLKGAINGLNPETIQYGAIDYALQ-----MKLISESDH 289  
 QY 282 NYMKEANOMPNGCQDLISTCKQTNRTALADYALCAEATNMCRDNVEGPPYAFAGGVYDI 341  
 DB 290 ESLK-----QDYVEQNTIKKCSLGGGLVCDASVEVCSIFENKIVAKKSGUNYDI 340  
 QY 342 RHP-----YDDPTPPSYNKFELAKDSVMDAIGV-NINYTQSNNDVYAFQOTGDFWPN 394  
 DB 341 RKKCVGSLCYDP---SRMEIFLNKENVRKALGVGDIFVSCSSIVY-----DAMIED 389  
 QY 395 FIEDLEEILALPVRLSYLIYGDADYICNWFEGGQAVSLAANYSOAQRSGYTPKLVNGVY 448  
 DB 390 WMONLE--VKPISLVNDGILNLYAGYDILICNWLGNRSWNSMEMSGKEAFVSSREBFTVD 447  
 QY 449 KVNVEGETREYGNFSTRVYAEAGHEVPYIOPIASLOLFNRTIFG 494  
 DB 448 LVDGKEAGLLKNHGHPLTELKYNAGHMVPMQPKVASLOLQNMWOG 493

## RESULT 9

Q9XH61 PRELIMINARY; PRT; 501 AA.  
 AC Q9XH61;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Serine carboxypeptidase.  
 OS Matricaria chamomilla.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;  
 OC Anthemideae; Matricaria.  
 ON NCBI\_TaxID=98504;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Kohchi C., Yasuda H., Hirata T.;  
 RA "Isolation of a cDNA encoding for a carboxypeptidase, having leucine  
 RA zipper structure at the N-terminal region, from the cultured shoot  
 RA primordia of Matricaria chamomilla";  
 RL Plant Biotechnol. 16:409-412(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Kohchi C., Yasuda H., Hirata T.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF141384; AAD42963.2;  
 DR HSSP; P00729; IYSC.  
 DR MEROPS; S10.009;  
 DR InterPro; IPR001563; Serine\_carbpept.  
 DR InterPro; IPR000379; Ser\_estrs\_site.  
 DR Pfam; PF00450; serine\_carbpept; 1.  
 DR PRINTS; PR00724; CRBOXYPTASEC.  
 DR ProDom; PD001189; Serine\_carbpept; 1.  
 DR PROSITE; PS00560; CARBOXYPEPT\_SER\_HIS; 1.  
 DR PROSITE; PS00131; CARBOXYPEPT\_SER\_SER; 1.  
 KW Carboxypeptidase.  
 SQ SEQUENCE 501 AA; 55973 MW; DEDFD8DB41880A66 CRC64;

Query Match 16.0%; Score 461; DB 10; Length 501;  
 Best Local Similarity 28.1%; Pred. No. 2.4e-24;  
 Matches 131; Conservative 76; Mismatches 188; Indels 72; Gaps 15;  
 QY 39 AEGVCETTPGVKSYSGY--VDTSPESHITFFWFFEARHNPETAPITLWNGPGSDSLIGL 96  
 DB 80 SDGV--TVEDLGHAGYQIQHSHAAKMFYFFESRNNKKD-PVWILTGGPGCSSLAL 136  
 QY 97 FEELGPGCHVNSTFDDYINPHSWNEVSNLLFLSQPLGVGFSDYTDVDSINPVTGVWNS 156  
 DB 137 FYENGPFKIADNMTLVWNEYGWDQASNLIVYDQPTGTGFSYSSDKDIRHDEQGV----- 191  
 QY 157 FAGVQGRYPTIDATLIDTTLAAEAWEILQGLSLSDRSVQSKDFSLWTSYGGHY 216  
 DB 192 -----SDDYDLQFAFFTEHPF-----VDNDFYITGESYAGHY 225

QY	217	GRAFFNHEFEQNERIANGSVNGVQLNFNSLGIINGIIDEAIOAPYYPFAVNNYTGKAV	276
Db	226	IFAIAARVHOGNK-----AKEGHIHLKGFICNGLTDPALQYQAYTDYALD-----MGLI	276
QY	277	NETVYNYMKFANPMGNCODLISTCKOTNR--TALADYALCAEATNMCNDRNVEGPFYAFAG	335
Db	277	KESQY---KRINLIVLPCAIAKLGTDGTGTVSCMAAYVVC-----NTIFFSILSIAG	325
QY	336	R-GVYDIRHP-----YDDETPPSYNNKFLAKDSVMDAIGV-NINTQSNNDVYXAFQOT	387
Db	326	NINXYDIRKECVTSMCYDF-----SDMETLLNKKSVRQALGVGDIEFVSCSTTVTTAMLVD	381
QY	388	GDFWPNFTEDLEILLALPVRSLIYGDADYICNWFQGVAVSLAANTSOAAQFSRAGYTP	447
Db	382	---LNRNLACIPLELLEDGKMLVYAGYDVICNWLGNSSVWVHAMENKGEQNALSEAP	438
QY	448	LKYNVEYGTREYGNFSFTRVYAGHEVYYPYQIASLQFNRTIFG	494
Db	439	FEVDGSEAGLLKSGPLSEFLKVDHAGHMVMDQPKAALEMLKRWMDG	485
RESULT 10			
Q8L6A7	Q8L6A7	PRELIMINARY;	508 AA.
AC	Q8L6A7	PRELIMINARY;	508 AA.
DT	01-OCT-2002 (Tremblrel. 22, Created)		
DT	01-OCT-2002 (Tremblrel. 22, Last sequence update)		
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)		
DE	Carboxypeptidase type III.		
GN	CP-III.		
OS	Theobroma cacao (Cacao) (Cocoa).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Malvales; Malvaceae; Byttnerioideae; Theobroma.		
OX	NCBI_TaxID=3641;		
RN	SEQUENCE FROM N.A.		
RA	Tazi H., McCarthy J., Bucheli P., Laloi M.;		
RA	"Molecular characterisation of a type-III carboxypeptidase (cacao Cp-		
RL	III) from Theobroma cacao seeds.";		
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AJ131514; CAC86383.1;		
DR	InterPro; IPR001563; Serine_carbpept.		
DR	pfam; PF00450; Serine_carbpept. 1.		
DR	PRINTS; PR00724; CRBOXYPTASEC.		
DR	ProDom; PD001189; Serine_carbpept; 1.		
DR	PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.		
DR	PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.		
SQ	SEQUENCE 508 AA; 56521 MW; 68B63707DC686BFE CRC64;		
Query Match			
Best Local Similarity 15.9%; Score 460; DB 10; Length 508;			
Matches 146; Conservative 88; Mismatches 206; Indels 96; Gaps 21;			
QY	1	LPGSTASVGRRQI-----PKNPTGVKTLTANNVTI-----RYKEPGAEGVCE	44
Db	34	LGGSEFSIHAKKLRLINLPKEVNV---VDGQVSLPDSRLVREKREFPNL-----	85
QY	45	TPPGVKS-----YSGY---VDTSPESHFTFFFAHNPETAPITLWNLGPGSDSLIGL	96
Db	86	AVPGGVSEDLGHAGYKLANSHDARMFYFFFSR-NSKKDPVVIWITGGPGCSSEAL	144
QY	97	FEELGPCVNSTFDDYINPHSWNEVSNLELISQPLGVGFSYSDTVDGSIINPTGVVENS	156
Db	145	EYENGFTTIAEMSLINWQYGDMAASNLIIYVQPIGTGFSYS-----SD	188
QY	157	FAGVQGYRTIDATLIDTTLNAAEAWFIQGLSPLDSRVQSKDFSLWTESYGGHY	216
Db	189	RRDIRHNEDEVSNDLYD-----FLQAFPAEHPEF-----EKNDFYITGESYAGHY	233
QY	217	GRAFFNHEFEQNERIANGSVNGVQLNFNSLGIINGIIDEAIOAPYYPFAVNNYTGKAV	276

```

DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00450; serine_carbpept; 1.
DR PRINTS: PR00724; CRBOXYPIASEC.
DR ProDom: PD001189; Serine_carbpept; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
SQ SEQUENCE 510 AA: 56896 MW; 134BC30F4E64BF0A CRC64;

Query Match 15.2%; Score 439; DB 10; Length 510;
Best Local Similarity 26.6%; Pred. No. 8.9e-23;
Matches 123; Conservative 81; Mismatches 189; Indels 69; Gaps 14;

QY 46 TPGVKSY---SGY--VDTSPESHITFFWFFFEARHNPEATITLWLNKGPGSDSLICLFEEL 100
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
93 SPSVQDGHGAGYKLPNKAARMEYFFESRTN-KADPVVIMLTGGPGCSSELALFYEN 151
QY 101 GPCVNSTFDDYINPHSNVEVSNLLFLSQPLGVGFSYDVTGDSINPVTGVVENSFPAGV 160
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
152 GPTVSNSSLSWNEFGWKASNLIVYDQPVGTGFSYTSQSDLRHDEGVDG----- 202
QY 161 QGRYPTIDATLIDTINLAARAAWEILOGLSLSDSRVQSKDFSLWTESYGGHYGPAF 220
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
203 -----SNDLYDFLOAFKHEP-----QFVKNDFYITGESYAGHYIPAL 240
QY 221 ENHYEQNERIANGSVNGVOLNFSGLIINGIIDEATQAPYYPEFAVNTYGIKAVNEIV 280
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
241 ASRVHRGNK-----NKEGTHINLKGFALGNLTNPEIQYGAYADYALDMNLITQSDHNL 295
QY 281 YNMYKANQMPNCCQDLISCTCKQTNRTALADYA---LCAEATNMCRDNVGPYVAFAGR- 336
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
296 NRY-----YATCQOSIKESADGEGDACASSYTV-CNNIFQKTMIDIAGNV 340
QY 337 GYVDIRHPYDDPTPPSYN--KFLAKDSVMDAIGV--NINTQSNNDVYVYAFQQTGDFWVP 393
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
341 NYDVRKQCEGSLCYDFSNNMENFNQKSVRKALGVGDIEFVSCSTAVYEAMQMD----WM 396
QY 394 NTE-ELEEITLAPVRVSLIYGDAIYICNFWGQAVSLAANYSQAAQFRSAGYIPLKVN 452
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
397 RNLEVGIPALLQDGIKLLVAGEYDLICNLWLGSKWVHEMWSQKEFEVAAATVPFVDN 456
QY 453 VEYGETREYGNFSTRVYEGAGHVPYQPIASQLFNRTIFG 494
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
457 KEAGLMKNYGSUTFLFKVDAGHVMVPMQPKAALQMLQNWMOG 498

RESULT 12
Q93ZC3 PRELIMINARY; PRT; 510 AA.
AC Q93ZC3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE AT3945010/F14D17.80.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H.A., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY057639; AAL15270.1; -.
DR InterPro: IPR001563; Serine_carbpept.
DR InterPro: IPR000379; Ser_estrs_site.

```

```

DR Pfam: PF00450; serine_carbpept; 1.
DR PRINTS: PR00724; CRBOXYPIASEC.
DR ProDom: PD001189; Serine_carbpept; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
SQ SEQUENCE 510 AA: 56837 MW; 704BC00C4E64BF08 CRC64;

Query Match 15.2%; Score 438; DB 10; Length 510;
Best Local Similarity 26.6%; Pred. No. 1e-22;
Matches 123; Conservative 80; Mismatches 190; Indels 69; Gaps 14;

QY 46 TPGVKSY---SGY--VDTSPESHITFFWFFFEARHNPEATITLWLNKGPGSDSLICLFEEL 100
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
93 SPSVQDGHGAGYKLPNKAARMEYFFESRTN-KADPVVIMLTGGPGCSSELALFYEN 151
QY 101 GPCVNSTFDDYINPHSNVEVSNLLFLSQPLGVGFSYDVTGDSINPVTGVVENSFPAGV 160
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
152 GPTVSNSSLSWNEFGWKASNLIVYDQPVGTGFSYTSQSDLRHDEGVDG----- 202
QY 161 QGRYPTIDATLIDTINLAARAAWEILOGLSLSDSRVQSKDFSLWTESYGGHYGPAF 220
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
203 -----SNDLYDFLOAFKHEP-----QFVKNDFYITGESYAGHYIPAL 240
QY 221 ENHYEQNERIANGSVNGVOLNFSGLIINGIIDEATQAPYYPEFAVNTYGIKAVNEIV 280
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
241 ASRVHRGNK-----NKEGTHINLKGFALGNLTNPEIQYGAYADYALDMNLITQSDHNL 295
QY 281 YNMYKANQMPNCCQDLISCTCKQTNRTALADYA---LCAEATNMCRDNVGPYVAFAGR- 336
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
296 NRY-----YATCQOSIKESADGEGDACASSYTV-CNNIFQKTMIDIAGNV 340
QY 337 GYVDIRHPYDDPTPPSYN--KFLAKDSVMDAIGV--NINTQSNNDVYVYAFQQTGDFWVP 393
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
341 NYDVRKQCEGSLCYDFSNNMENFNQKSVRKALGVGDIEFVSCSTAVYEAMQMD----WM 396
QY 394 NTE-ELEEITLAPVRVSLIYGDAIYICNFWGQAVSLAANYSQAAQFRSAGYIPLKVN 452
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
397 RNLEVGIPALLQDGIKLLVAGEYDLICNLWLGSKWVHEMWSQKEFEVAAATVPFVGN 456
QY 453 VEYGETREYGNFSTRVYEGAGHVPYQPIASQLFNRTIFG 494
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
457 KEAGLMKNYGSUTFLFKVDAGHVMVPMQPKAALQMLQNWMOG 498

RESULT 13
Q17679 PRELIMINARY; PRT; 2105 AA.
AC Q17679;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE X16B4A.2 protein.
GN X16B4A.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Thomas K.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z70203; CAA94110.1; -.
DR EMBL; AL023825; CAA94110.1; JOINED.
DR EMBL; Z93339; CAA94110.1; JOINED.
DR EMBL; Z93339; CAB07544.1; -.
DR EMBL; AL023825; CAB07544.1; JOINED.
DR EMBL; Z70203; CAB07544.1; JOINED.
DR EMBL; AL023825; CAA19443.1; -.
DR EMBL; Z93339; CAA19443.1; JOINED.
DR EMBL; Z70203; CAA19443.1; JOINED.
DR HSSP; P10619; LIVV.
DR WormPep; Y16B4A.2; CE21374.
DR InterPro: IPR001563; Serine_carbpept.

```

```
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00450; serine_carbpept. 4.
DR PRINTS: PR00724; CRBOXYPTASEC.
DR ProDom: PD001189; Serine_carbpept. 6.
DR PROSITE: PS00560; CARBOXYPEPT_SER_HIS; 4.
DR PROSITE: PS00131; CARBOXYPEPT_SER_SER; 2.
SQ SEQUENCE 2105 AA; 234335 MW; 139CE6E96D75B3FB CRC64;

Query Match 14.3%; Score 413.5; DB 5; Length 2105;
Best Local Similarity 24.6%; Pred. No. 4.4e-20; Indels 173; Gaps 26;
Matches 150; Conservative 85; Mismatches 202;

QY 4 STPASGRRLPKNPQVTKLTATANNVTIRYKPGAEGVCTTPGV-----KSYSGYVD 57
DB 1521 STPAGI-----DVTPOPLNIVISQTNCT-----TGQTDRLINLGLPADMLFKQYSGFLD 1570
QY 58 TSPESHTFFWFEARINPETAPITIMLNGPGSDSLGLFEEELGPGCHVNSTEDDYI-NPH 116
DB 1571 GLSGHKVHYWLVSENNPTDPLLLWNGPGSSSLMGLFEENGPRFVSKDSQTLNRNPY 1630
QY 117 SNNEVSNLLFSLQPLGVGFYSYDVTGDSINPVIGVVENSSFAGVQGRYPTIDATLIDTNN 176
DB 1631 SNKFNVLVLSPIGVGSYA-----YNNINLQYDDVT- 1664
QY 177 LAEAAWEITLQGLSLPSLDSRVQSKDFSLWTSYGGHYGPAFFNHFYEQNERIANGSV 236
DB 1665 -TAQENYAALKSFPAQY-----QYTTSDFTTGESYAGVYLPGL-----SALLVOGK 1712
QY 237 NG-VQINFNLSGLIINGLIDE-----AIQAPYYPEFAVNTYTGKAVNEIYVN----- 282
DB 1713 SGDININIKGVSGVNGVIDKRDIDMNSQLHYQY-----HG--GISATYNTALQLOCC 1762
QY 283 -----YMKFANOMPNG-----COD-LISTCKQTNRTALADYALCAEATNMCRD 324
DB 1763 SGDEFKCRSDRTNFTNNSIPWGLSDPCYDFVVAATCAQLLLTAFDPYNNYQOCWTI--- 1819
QY 325 NVEGPHYAFAGRGVYDIRHPYDDPTPPSYNKK-----FL 358
DB 1820 -----PYNDTPRTPYGETWTGTYNIESDFLNGPCYDSDSAMESYL 1860
QY 359 AKDSVMDAIGV--NINQTSNDVYAFQGTGFVWFNFTEDEEILIA--LPRVVSLEY 413
DB 1861 NRPVVRKALIPDSVPYWAANNINLIAYNOQVDSIPFN-----LQIIMANAPAFKMLLYS 1916
QY 414 GDADYICNWRFGG-----AVSLAANYSQAA-----QFRSAGYTPKLVNGVVEYGETREYG 462
DB 1917 GDADTYNWLGAELFTANFAALGLTSSPRAQWTFQYNST-FQP-TVAGYQISYTSNAI 1974
QY 463 NFSFTRVYEHAGHYVYPTIASQLFNRTIFGWDIAEGQKKIWPSTKINGTATATHTQSS 522
DB 1975 NIDVLTVKSGHFVPLDRPOQALQM-----IYNFVKSRGYNTPFDLNSFTTTTITSTTT 2030
QY 523 VP--LPTATS 530
DB 2031 TPGTGPTVTA 2040

RESULT 14
Q9FMX9 PRELIMINARY; PRT; 499 AA.
AC Q9FMX9 PRELIMINARY; PRT; 490 AA.
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Serine carboxypeptidase II-like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
```

```
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned pl clones."
RL DNA Res. 4:401-414(1997).
DR EMBL: AB007648; BAB11176.1; -.
DR HSSP: P08819; 1WHT.
DR MEROPS: S10.005; -.
DR InterPro: IPR001563; Serine_carbpept.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00450; serine_carbpept. 1.
DR PRINTS: PR00724; CRBOXYPTASEC.
DR ProDom: PD001189; Serine_carbpept. 2.
DR PROSITE: PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE: PS00131; CARBOXYPEPT_SER_SER; 1.
DR Carboxypeptidase.
KW Carboxypeptidase.
SQ SEQUENCE 499 AA; 56329 MW; 6150FB81A2E7BCBD CRC64;

Query Match 14.1%; Score 406.5; DB 10; Length 499;
Best Local Similarity 26.6%; Pred. No. 1.7e-20;
Matches 137; Conservative 75; Mismatches 196; Indels 107; Gaps 22;

QY 33 RYKE-PGAEVGVCTTPGVK--SYSGYVDTSPESH---TFWFEARINPETAPITLWNG 86
DB 51 RVKELPG-----QPVKFRQAGYV-TVNETHGRALFYWFTEATQNPSSKKPVLLWNG 102
QY 87 GPQSDSI-IGLEELGCHVNSTFDD--YINPHSNWVSNLLFSLQPLGVGFYSYDVTVDG 143
DB 103 GPGSSIGFGAAEELGFFPQNSQPKLKNPFSWNKAANLLFLESPGVGFSYNTI--- 159
QY 144 SINPVTGVVSNSSFAGVQGRYPTIDATLIDTNNLAEEAAWEITLQGLSLPSLDSRVQSK 203
DB 160 -----SRDIKOLGDT-VTARDSYNLFVNWFKREF---QYKSH 192
QY 204 DFSLWTSYGGHYGPAFFNHFYEQNERIANCSVNGVQINFNLSGLIINGLIDEAQAQYYP 263
DB 193 DFYTAGESYAGHYVPQSELITYKENKIAASKDF-----INLGLMIGNALLDDETDQKGM 248
QY 264 EFAYNNYTGKAVNETVYNYMKFANQM-PNSQCDLIS-----TCROTN 305
DB 249 EYAWDHAVISDALYEKYNKNCDFQKLVTKCNDALEDFDVYKILDMYSLYAFKCVPTS 308
QY 306 RTALADYALCAEATNMCNDNVGPPYAFAGRGVYDIRHPYDDPTPPSYNKFELAKDSVMD 365
DB 309 TNSSTSHSVAGNRPLPAFRSILRPLRILSHNEGWRMAAGY-DPCASEYTEKYMNRKDQVE 367
QY 366 AIGVNYNTOSNDVYAFQGTGFV--W-----PNFIEDLEELALPVRVSLIYGDADYI 419
DB 368 ALHANV-----TNISYPWTHCSDTVSWSDAPASMLPTLRLTSAGLRVWVFGSDTD-- 419
QY 420 CNWFGGOAVSLAANY-----QAAQFRSAGYTPKLVNGVVEYGETREYGNFSFTRVYEA 473
DB 420 -----GRIPVTATRYSLKGLKLVQDWTWPYTKLOVG-----GWTVEYDGLMFVIRGAG 470
QY 474 HEVYYPOTIASQLFNRTIFGWDIAEGQKKIWPSTY 508
DB 471 HQVPTFKPREALQIHHFL-----GNKKL-PTF 497

RESULT 15
Q9FYP7 PRELIMINARY; PRT; 490 AA.
AC Q9FYP7 PRELIMINARY; PRT; 490 AA.
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to Hordeum vulgare carboxypeptidase D (Putative
DE carboxypeptidase D).
GN P0011G08.38.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```



Search completed: September 16, 2003, 22:51:43  
Job time : 80 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 16, 2003, 22:52:38 ; Search time 6324 Seconds  
(without alignments)  
3473.825 Million cell updates/sec

Title: US-09-712-338-2\_COPY\_19\_555  
Perfect score: 2887  
Sequence: 1 LFGSTPASVGRRLPKNPQTG.....HTQSSVPLPTATSMSSVGMGA 537

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-Q=/cgn2\_1/USPTO.spool.p/US09712338/runat\_16092003\_144914\_19130/app\_query.fasta\_1.711  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09712338 @CGN\_1.1.4958 @runat\_16092003\_144914\_19130 -NCPU=6 -ICPU=3  
-NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl: \*  
1: gb.ba: \*  
2: gb.htg: \*  
3: gb.in: \*  
4: gb.om: \*  
5: gb.ov: \*  
6: gb.pat: \*  
7: gb.ph: \*  
8: gb.pl: \*  
9: gb.pr: \*  
10: gb.ro: \*  
11: gb.sts: \*  
12: gb.sy: \*  
13: gb.un: \*  
14: gb.vi: \*  
15: em.ba: \*  
16: em.fun: \*  
17: em.hum: \*  
18: em.in: \*  
19: em.mu: \*  
20: em.om: \*  
21: em.or: \*  
22: em.ov: \*  
23: em.pat: \*  
24: em.ph: \*  
25: em.pl: \*  
26: em.ro: \*  
27: em.sts: \*  
28: em.un: \*

29: em.vi: \*  
30: em.htg\_hum: \*  
31: em.htg\_inv: \*  
32: em.htg\_other: \*  
33: em.htg\_mus: \*  
34: em.htg\_pla: \*  
35: em.htg\_rod: \*  
36: em.htg\_mam: \*  
37: em.htg\_vrt: \*  
38: em.sy: \*  
39: em.htgo\_hum: \*  
40: em.htgo\_mus: \*  
41: em.htgo\_other: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2883	99.9	1662	6	AR129928 Sequence
2	2526	87.5	2245	8	AF394242 Aspergill
3	2338	81.0	1656	6	AX534871 Sequence
4	1983	68.7	3150	6	AX534814 Sequence
5	980.5	34.0	1872	6	AX534872 Sequence
6	977	33.8	3221	6	AX534815 Sequence
7	941	32.6	1665	6	AX534866 Sequence
8	916.5	31.7	1581	6	AX534824 Sequence
9	894	31.0	2940	6	AX534767 Sequence
10	887.5	30.7	3080	6	AX534809 Sequence
11	652.5	22.6	4308	8	D86560 Schizosacch
12	652.5	22.6	37000	8	SPAC119G12
13	636	22.0	1527	6	AX594884 Sequence
14	636	22.0	2016	8	SCYBR139W
15	636	22.0	2027	6	AX536454 Sequence
16	636	22.0	29686	8	SC29711
17	636	22.0	50277	2	AC138524_5
18	630	21.8	1611	6	AX534826 Sequence
19	623	21.6	2002	6	I33983 Sequence 3
20	623	21.6	2002	6	I74375 Sequence 3
21	623	21.6	2002	6	I77239 Sequence 3
22	596.5	20.7	2509	8	AF085063 Pichia an
23	596	20.6	1985	8	YSACARPEPY
24	595.5	20.6	2214	8	PAU67174
25	594.5	20.6	3850	6	E12103
26	594.5	20.6	3850	8	PPPRC1GEN
27	594	20.6	2068	6	I33982
28	594	20.6	2068	6	I74374
29	594	20.6	2068	6	I77238
30	580	20.1	2660	6	AX534769 Sequence 1
31	568.5	19.7	2503	6	A75535 Sequence 2
32	568.5	19.7	2503	6	I70282 Sequence 3
33	548.5	19.0	2280	8	AB051820 Aspergill
34	541.5	18.8	1653	6	AX536918 Sequence
35	527.5	18.3	2632	8	YSCPRCCPY
36	527.5	18.3	15581	8	SC8175
37	523.5	18.1	2632	6	AR129959 Sequence
38	479	16.6	1787	8	AY072822 Gossypium
39	476.5	16.5	1772	8	RICCBP31
40	474	16.4	1551	6	AX412745 Sequence
41	474	16.4	1551	6	AX505988 Sequence
42	474	16.4	1551	8	AY149954 Arabidops
43	474	16.4	1771	8	AY091767 Arabidops
44	470.5	16.3	3298	8	ABGSCFZ
45	461	16.0	1816	8	AF141384 Matricari

ALIGNMENTS

RESULT 1

AR129928  
LOCUS AR129928 1662 bp DNA linear PAT 16-MAY-2000  
DEFINITION Sequence 1 from patent US 6187578.  
ACCESSION AR129928  
VERSION AR129928.1 GI:14117825  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1662)  
AUTHORS Blinkovsky,A., Berka,R., Rey,M., Golightly,E., Klotz,A., Mathisen,T.Erik., Dammann,C. and Brown,K.M.  
TITLE Carboxypeptidases and nucleic acids encoding the same  
JOURNAL Patent: US 6187578-A 1 13-FEB-2001;  
FEATURES Location/Qualifiers  
source 1..1662  
BASE COUNT 396 a 468 c 398 g 400 t  
ORIGIN /organism="unknown"  
Alignment Scores:  
Pred. No.: 7.36e-218 Length: 1662  
Score: 2883.00 Matches: 536  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.86% Indels: 0  
DB: 6 Gaps: 0  
US-09-712-338-2\_COPY\_19\_555 (1-537) x AR129928 (1-1662)  
Qy 1 LeuProGlySerThrProAlaSerValcIyArgArgGlnLeuProLysAsnProThrGly 20  
Db 55 CTTCCAGGAAGTACACCGCGTCGTCGGTAGAAGACAGCTACCCAGAACCCACCGCG 114  
Qy 21 VallysThrLeuThrThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGlu 40  
Db 115 GTCAAGACTCTTACACCGCAACCAATGTCACCTCCGGTACAAAGAACCCGGGCGCAG 174  
Qy 41 GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 60  
Db 175 GCGCTCCGAGACTACCCGGGTGTCAAACTCTACTCTGGATATGTCACACCTCTCCC 234  
Qy 61 GluSerHisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAlaProIle 80  
Db 235 GAGTCCCATACCTCTCTGGTCTCTTCGAAGCCAGACATAACCAGAAACTGCACCTATC 294  
Qy 81 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeu 100  
Db 295 ACATGTGGTTCAAATGGTGGCCCGGAAGCCGATCTTTGATCGGTCCTTCGAAGAGTTG 354  
Qy 101 GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGlu 120  
Db 355 GGCCCTTGCCATGTCAATTCGACTTTTGATGACTACATCAACCCCTCACTCGTGGAAACG 414  
Qy 121 ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 140  
Db 415 GTCTCCAAATTTACTATTCCTGTGCCAGCATTTGGAGTCGGCTTTTCATATAGTGATAG 474  
Qy 141 ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal 160  
Db 475 GTTGATGGGTCCATTAACCTGTAACTGGGGTCGTGCAAAATTCGAGCTTTTCAGGAGTT 534  
Qy 161 GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGlu 180  
Db 535 CAGGCCCGGTACCCAACCATGTATGCCACTCTGTATCGATACCTACCAAACTTTCGCCAGAG 594  
Qy 181 AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal 200  
Db 595 GCCCGTTGGGAGATCTCTCAAGAGATTCTCTATGGGACTACCTTAGCTTAGGCTAGGGTG 654  
Qy 201 GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 220  
Db 655 CAGTCTAAGGACTCTCAGTCTATGGACCGAGAGCTATGGAGGCACTATGGTCTCTGCATTC 714

SOURCE Aspergillus oryzae  
ORGANISM Aspergillus oryzae  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
REFERENCE 1 (bases 1 to 2245)  
AUTHORS van den Broek, P.  
TITLE Direct Submission  
JOURNAL Submitted (22-JUN-2001) Bioscience, Nestec S.A., P.O. Box 44,  
Lausanne CH-1000, Switzerland  
FEATURES  
source 1..2245  
Location/Qualifiers  
/organism="Aspergillus oryzae"  
/mol\_type="genomic DNA"  
/strain="Tk3"  
/db\_xref="taxon:5062"  
1..2245  
/gene="cpi"  
join(1..349,409..513,576..692,743..866,941..956,  
1008..1124,1187..1237,1296..1449,1522..1580,1670..2245)  
/gene="cpi"  
/product="serine carboxypeptidase 1"  
join(1..349,409..513,576..692,743..866,941..956,  
1008..1124,1187..1237,1296..1449,1522..1580,1670..2245)  
/gene="cpi"  
/EC\_number="3.4.16.6"  
/codon\_start=1  
/product="carboxypeptidase S1"  
/protein\_id="AAK7166.1"  
/db\_xref="GI:15004616"  
/translation="MRGEFLSVLPVVAASWALPGSTFASVGRRLKPNPTGVKLTIT  
ANNVTRYKEFGAEVCTTPGVKSYGYVDIPESTIFFWFPEARNPTAPITLW  
NGGPGSLGLGAEELGPCVHNSTFDYINPHSWNEVSNLLELSQPLGVGSYNDIVD  
GSTNPYTVGNSFAGVGKRYPTIDATLIDTTLNAAEAWEILLOGLSLPLSDSRV  
QSKDFSLWTSYGHYGPAPFHEFYEQNERIANGSVNGVNLNENGLINGIDEAIO  
APYPEFANVTYGIKAVNETVYIMKFAFMNGCCQDLISTCKQNRNALADYALCA  
EAINMCKQVEGYPYAFAGVGYDIRHPYDDPTPPSYNKFLEADSYMDAIGVNIYNT  
QSNDDYVAFQGTDFWPNFIEDLEILALPVRSLYGDADYICNWFQGAVALAA  
NYSQAQFRSAGYPLKVGVEYGETREYGNFSFTRVYEAGHEVPYPTIASLQLEFR  
TIFGWDIAEQKKIWPYSYKINGTATATHTQSSVPLPTATSMSSVGN"  
sig\_peptide 1..54  
BASE COUNT 548 a 603 c 524 g 570 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.43e-189 Length: 2245  
Score: 2526.00 Matches: 530  
Percent Similarity: 72.54% Conservative: 1  
Best Local Similarity: 72.40% Mismatches: 6  
Query Match: 87.50% Indels: 201  
DB: 8 Gaps: 9  
US-09-712-338-2\_COPY\_19\_555 (1-537) x AF394242 (1-2245)  
QY 1 LeuProGlySerThrProLysValGlyArgGlnLeuProLysAsnProThrGly 20  
|||||  
Db 55 CTCCAGGAAGTACACCGCGTCCGCTCGGTAGAGACAGCTACCCAGAACCCACCGGG 114  
QY 21 ValLysThrLeuThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGlu 40  
|||||  
Db 115 GTCAAGACTCTTACAAACCGCAACAATGTCCACATCCGGTACAAAGAACCGGGCAGAG 174  
QY 41 GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 60  
|||||  
Db 175 GGGCTCGGAGACTACCCGGGTGCAAAATCTCTCTGGATATGCGACACTCTCC 234  
QY 61 GluSerHisThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProIle 80  
Db 235 GAGTCCCATACCTCTCTCTGTTCTTCGAAGCCAGACATAACCCAGAACTGCACCAATC 294  
QY 81 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGlu----- 98  
|||||  
Db 295 ACATTGTGTTGAATSGTGCCCTGGAAGCGATTCTTTGATCGGTCTCTCGA-AGTTT 353

QY 99 -----GluLeu 100  
354 GCTTACGCTGTTACAAAGTANGCTCTTTTAGGTCTCGGTTAACCTTGGTTAGAGTGG 413  
Db  
QY 101 GlyProCysHisValAsnSerThrPheAspTyrIleAsnProHisSerTrpAsnGlu 120  
|||||  
Db 414 GSCCCTTCCCATGTCATTCGACTTTTGTATGACTACATCAACCCCTCACTCGTGGACGAG 473  
QY 121 ValSerAsnLeuLeuPheLeuSerGlnProLeuGly-----132  
474 GTCCTCAATTTACTATTCTCTGTCCAGCCATT-GGGAGTCGGTATGGATTGCCGCCACCT 532  
QY 133 -----ValGlyPheSerTyrSerAsp 139  
533 CCCCTACTGGGTCCTCCCATATTGACGAGGTGCTTCCCGTAGGCTTTTCATATAATGAT 592  
QY 140 ThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGly 159  
|||||  
Db 593 ACGGTTGATGGGTCCATTAACCCCTGTAACCTGGGTCTGCGAAATTCGAGCTTTGCGAG 652  
QY 160 ValGlnGlyArgTyrProThrIleAspAlaThrLeuIle-----172  
653 GTTCAGGGCCGCTACCCCAACCATTCATGCCACTCIGAT-CGGTAACGGTTTCTCCGGTC 711  
QY 173 -----AspThrThrAsnLeuAlaAlaGluAla 182  
712 CACATGCTCTTATTACTGATACGATAGCAGATAC1ACCAATCTTCCGCGACAGAGCCGCT 771  
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202  
|||||  
Db 772 TGGGAGATCCCTGCAAGGATTCCTTAGTGACTACTAGCTTGGACTCTAGGCTGCACTGCT 831  
QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyGly-----214  
832 AAGGACTTCAGTCTATGAGCGAGAGCTAIGAGG- GTAGTGGTGTCCCTCCTCGGAATC 890  
QY 215 -----HisTyrGly 217  
891 CTTAAACGGCGGTGATGGTTCGCTGATAGGTTCCTTCCTTCGAAAGGCATATGTT 950  
QY 218 ProAla-----Phe 220  
951 CCTGCAGTGTCTTACCTGGACCATGTCTCAAAGAAAGCAAGACTAACCAAGACCTAGTTC 1010  
QY 221 PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 240  
1011 TICAATCAITTTACGAGCAGAAAGAGAGATTCGCCAAGCGTAGTGTITATGGTGTTCAG 1070  
QY 241 LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAla--- 259  
1071 CTTAAATTCAACTCTCTGGGAATTAITACGGCATCAICGACGAGGCGATCCAGGT-ATA 1129  
QY 259 -----259  
1130 CCCTCGAAGACGCCCATAGAACATCTCCGAACAAAGCTCTAACGTATGCCCAATAGGCC 1189  
QY 260 ProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAla-----275  
1190 CCTTACTACCTGAATTCGCTGTGACATACCTACCTACGTTATCAAGGCTGTAAAGTTCC 1249  
QY 276 -----ValAsnGluThrVa 280  
1250 TTAAGCGCATGAGATTATATCTCTGATACCTCATCACCATGAAAGSICAAACGAGACCGT 1309  
QY 280 lTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGluAspLeuIleSerTh 300  
1310 CTACAACTCATAGAACTTTGCCCAACCAAAATGCCAAATGGTTGCCAGGATTGATTCCAC 1369  
QY 300 rCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAs 320  
1370 CTGCAACAGACAGAAACCGCACCCGCAITAGCTGACTAGCCCTCTGCGCCGAAGCCACCAA 1429

QY 320 nMetCysArgAspAsnVal----- 326  
 Db 1430 CATGTGCGGACAAATGTTGGTACGTTCTTCCCTCCCTTGGAAACAGGTAATGATT 1489  
 QY 327 -----GluGlyProTyrTyrAlaPheAlaGlyAr 336  
 Db 1490 CAAAATAGTGGGACAAATAACAGAAAATAGAGGGCCCATACTACGCCCTTCTGCTGGTGG 1549  
 QY 336 gGlyValTyrAspIleArgHisProTyrAsp----- 346  
 Db 1550 TGGTGTGTATGATATTTCGGCATCCATATGA-TGTAAGTAGAGCTTTTACTTCTATTATTACA 1608  
 QY 346 ----- 346  
 Db 1609 CACCGCATGCAATGATGCTGTTTCATTTGATGCAATAAATCCCCCTAATATGGTATTA 1668  
 QY 347 -AspProThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAl 366  
 Db 1669 GGACCGGACTCCCGCAAGTTATTACACAAATTTCTGGCCAGGAGCTCTGTCATGGAGCG 1728  
 QY 366 alieGlyValAsnIleAsnTyrThrGlnSerAsnAspValTyrTyrAlaPheGlnG1 386  
 Db 1729 TATCGGGGTCAACATCAACTACACCCCAATCAATATGACGTCTACACGCTTTCCAGCA 1788  
 QY 386 nThrGlyAspPheValIrrProAsnPheIleGluAspLeuGluIleLeuAlaLeuPr 406  
 Db 1789 AACAGGGGACTTGTCTGGCCCACTTCATCGAAGACCTCGAGGAGATCCTTGCTCTCC 1848  
 QY 406 oValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyG1 426  
 Db 1849 CGTGGGTCTCTCCCTCATCTATGGCGACGCGATTCATCTGCAACTGGTTCGGCGTCA 1908  
 QY 426 nAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGlyTyrTh 446  
 Db 1909 GGCGGTTCCTCCTCGCAACTACTCCCAAGCGCCCGAGTTCGCAAGCGCAGGGTACAC 1968  
 QY 446 rProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPh 466  
 Db 1969 GCCCCTGAAAGTCAACGGCGTCGAGTATGGGGAACCTCGCGAGTATGATGTAATTTCTCCTT 2028  
 QY 466 eThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuG1 486  
 Db 2029 CACTCGGGTCTATCAGGACGCGCATGAGTCCCATACTACCGCCCAATCGCCTCCCTGCA 2088  
 QY 486 nLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLysIleTyrPr 506  
 Db 2089 ATTGTTTAACGGGACTATCTCGGTGGATATCGAGAGGCGCCAGAGAAGATCTGGCC 2148  
 QY 506 oSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSerValProLeuPr 526  
 Db 2149 CAGCTACAAGACGAATGAGAGCGGTACAGCTACGCATACAGAGTCTGCTCGCGCTGCC 2208  
 QY 526 oThrAlaThrSerMetSerSerValGlyMetAla 537  
 Db 2209 TAGGCTACCAGCATGTCAGTGTGGTATGGCT 2242

## RESULT 3

AX534871  
 LOCUS AX534871 1656 bp DNA linear PAT 22-NOV-2002  
 DEFINITION Sequence 109 from Patent WO02068623.

ACCESSION AX534871

VERSION AX534871.1 GI:25261462

## KEYWORDS

Aspergillus niger

## ORGANISM

Aspergillus niger  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Eurotiiales; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.

## REFERENCE

1  
 Edens L., dijk Van A.A., Krubasik P., Albermann K., Stock A.,  
 Kimpel E., Klugbauer S., Wagner C., Fritz A., gustedt Von W.,  
 Heinrich O., Maier D., Spreafico F., Folkers U., Hopper S.,  
 Kemmer W., Tan P., Stiebler J. and Albarg R.  
 Novel genes encoding novel proteolytic enzymes

## TITLE

JOURNAL Patent: WO 02068623-A 109 06-SEP-2002;

DSM N.V. (NL)

FEATURES Location/Qualifiers

source

1..1656

/organism="Aspergillus niger"

/mol\_type="genomic DNA"

/db\_xref="taxon:5061"

BASE COUNT 411 a 426 c 409 g 410 t

ORIGIN

Alignment Scores: 6.37e-175 Length: 1656

Pred. No.: 2338.00 Matches: 419

Score: 88.87% Conservative: 52

Best Local Similarity: 79.06% Mismatches: 59

Query Match: 80.98% Indels: 0

DB: 6 Gaps: 0

US-09-712-338-2\_COPY\_19\_555 (1-537) x AX534871 (1-1656)

QY 1 LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly 20  
 Db 55 ATGCCCGAGATGAATGGTCACTACGATAAGAAGGAGGTACCAAAAGCGTCACATGGC 114  
 QY 21 VallysThrLeuThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGlu 40  
 Db 115 GTCAATTCGATAAAACCCCAACATGCTACTATCAGGTATAAGAACCAACGAGAACCGAA 174  
 QY 41 GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 60  
 Db 175 GGAATTTGTGACACACACCTGGGTCAATATCATCTCCGATATGCGATCTTCGCCA 234  
 QY 61 GluSerHisThrPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIle 80  
 Db 235 GAGTCGCATACTTCTTTTGGTTTTCGAGTCAACGCGTACCCCGAAATGATCCAGTG 294  
 QY 81 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeu 100  
 Db 295 ACTCTGGTGGTGAATGGTGGCCCTGGAAGCGATCTCTTGATGGGCTTTTGAAGAGTGG 354  
 QY 101 GlyProCysHisValAsnSerThrPheAspTyrIleAsnProHisSerTrpAsnGlu 120  
 Db 355 GGTCCGTGCATACACACAGTACGAGTACGAATCAATCAATCACTACTCTCTGGAACGAG 414  
 QY 121 ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 140  
 Db 415 GTCACCAATCTCTTTTCTGTCAGCCCTCGGTGGGTCTCTTACAGTGAACGC 474  
 QY 141 ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal 160  
 Db 475 GAGCGCGGTCTCTTGAAATCCATTTACTGGAGCCGTCGAGAGCGCTCTCTTCTGCTGAGT 534  
 QY 161 GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrAsnLeuAlaGlu 180  
 Db 535 CAGGTGCGATACCCAGTTATTGATGCCACTATCATCGACACGACGATATCGCTCACGC 594  
 QY 181 AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal 200  
 Db 595 GCAACCTGGGAGTGGCTTCAGGCTCTCCCTACGTGGCTCTCGACAGTAGATTCCGAAGTC 654  
 QY 201 GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 220  
 Db 655 AAGTCCAGGAGTTCAACTGCTGGACAGAGTACGGAGGACACTATGACGACGCTTC 714  
 QY 221 PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 240  
 Db 715 TTCAATCATTTCTACGAGCAAAATTCGAAGATCGCTAGCGGGGAGTCAATGCGCTCAA 774  
 QY 241 LeuAsnPheAsnSerLeuGlyIleLeuAsnGlyIleIleAspGluAlaIleGlnAlaPro 260  
 Db 775 CTGAATTTTAACTCCCTCGGATTAACAACGGCATCATGATGCGCGGATTCAGGACGAC 834  
 QY 261 TyrTyrProGluPheAlaValAsnAsnThrTyrTyrIleIleAlaValAsnGluThrVal 280

```

Db      835  TACTACCGAGACTTGGCGTTAAATAATACATATGCAATCAAGCTGTCAATGACACAGT 894
Qy      281  TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr 300
Db      895  TACAACATATATGAGTTCGCGAACACAGATGCCAATGATGATGCCAGGATGAGTTCGTCG 954
Qy      301  CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn 320
Db      955  IGTAATTTGACCAATAGACCTGCTTCTGATTATGCTATATGTATACAGAGACCCCAT 1014
Qy      321  MetCysArgAspAsnValGluGlyProTyrThrAlaPheAlaGlyArgGlyValTyrAsp 340
Db      1015  ATGTCAGAGGACAAATGTCGAGAGGCCCTTACTACCAAGTTTGGCGCGTGTATGAT 1074
Qy      341  IleArgHisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys 360
Db      1075  ATTCGGGACCCCTACATATGATGACCCCGCGCTTCTTGTGTTGTTGTTGTTGTTGTTG 1134
Qy      361  AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal 380
Db      1135  GACTCAGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1194
Qy      381  TyrTyrAlaPheGlnThrGlyAspPheValTyrProAsnPheIleGluAspLeuGlu 400
Db      1195  TATATGCAATTCAGACAGACCGCGACTTGTATGCGCAATTTCAATTGAGGACCTCGAA 1254
Qy      401  GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys 420
Db      1255  GAGATCCTCAACTCCCGTACCGGTCGTTGATGATGATGATGATGATGATGATGATGATG 1314
Qy      421  AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGlnPhe 440
Db      1315  AACTGTTCGCGGTGAGGCCATCTCAGTGCAGTTAACTACCCGATGCGAGCTCAGTTC 1374
Qy      441  ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu 460
Db      1375  CGTGCAGCGGATACACACCATGACAGATAGATGGGTGCAATACGGTGAGACTCGCGAG 1434
Qy      461  TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln 480
Db      1435  TATGCAACTTTCGTTCACCGCGTATATACGAGTGGGACGAGGTTCATCTACTATCAA 1494
Qy      481  ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGly 500
Db      1495  CCGATCGCAGCGTTCAGCTGTCAACCGTACTTTATTTGGATGGGATATGCGCGGT 1554
Qy      501  GlnLysLysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGln 520
Db      1555  ACAACTCAGATTGGCCGGAATAGCACCACGGGACATCGCAGGCTACACACAGGAG 1614
Qy      521  SerSerValProLeuProThrAlaThrSer 530
Db      1615  TCGTCGTGCCACTGTCCACGGCGTCACT 1644

```

## RESULT 4

AX534814

LOCUS

AX534814

DEFINITION

Sequence 52 from Patent WO02068623.

ACCESSION

AX534814

VERSION

AX534814.1

KEYWORDS

GI:25261343

SOURCE

Aspergillus niger

ORGANISM

Aspergillus niger

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

REFERENCE

1

AUTHORS

Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock, A.,

Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W.,

Heinrich, O., Maier, D., Spreafico, F., Folkers, J., Hopper, S.,

Kemmer, W., Tan, P., Stiebler, J. and Albarg, R.

Novel genes encoding novel proteolytic enzymes

Patent: WO 02068623-A 52 06-SEP-2002;

## FEATURES

Location/Qualifiers

source

/organism="Aspergillus niger"

/mol\_type="genomic DNA"

/db\_xref="taxon:5061"

BASE COUNT 801 a 778 c 738 g 833 t

## ORIGIN

Alignment Scores:

Pred. No.:	1-32e-146	Length:	3150
Score:	1983.00	Matches:	412
Percent Similarity:	65.07%	Mismatches:	50
Best Local Similarity:	58.03%	Indels:	185
Query Match:	68.69%	Gaps:	8

US-09-712-338-2\_COPY\_19\_555 (1-537) x AX534814 (1-3150)

Qy 1 LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly 20  
 Db 629 ATGCCCGAGAAATGATGTCATCTACGATAAGAGGAGTACCAAAAGCGTCCACTGCG 688  
 Qy 21 ValLysThrLeuThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGlu 40  
 Db 689 GTCAAATCGATAAAACCCCAACATGTCATCTAGTATAAGAACCCAGAACCGAA 748  
 Qy 41 GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 60  
 Db 749 GGAATTTGTGAGACACACCTGGGTCAATCACTCCGATATGTCGATCTTCGCCA 808  
 Qy 61 GluSerHisThrPheThrPhePheGluAlaAlaArgHisAsnProGluThrAlaProIle 80  
 Db 809 GAGTCGCATATCTTTCTTTTGGTTCAGTACCGCGGACCGGACCGGATGATCCAGT 868  
 Qy 81 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGlu 98  
 Db 869 ACTCTGGCTGATGTTGGCCCTGGAGCGATTCCTGTGATTTGGCTTTTGA-AGGTTG 927  
 Qy 98 ----- 98  
 Db 928 GCCAAATATCTCTGACGGGAAAGATAAATTCAGTTCATGCTGACGCTTCACAACA 987  
 Qy 99 GluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrp 118  
 Db 988 GAGTTGGTCCGTGTCACATCACACAGAGTAGAGAAATCAATCATCATCATCTCTCTG 1047  
 Qy 119 AsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyVal 133  
 Db 1048 AACGAGTCAACCAATCTCTTTTCTTCTCAGCCCTCGGTGTGGGTATGGGAATATG 1107  
 Qy 134 ----- 134  
 Db 1108 TGCCITCATACATCTCTGAGTACATTCCTTACGGTCTTATCIGGAGGGTTCCTTACAG 1167  
 Qy 138 rAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAl 158  
 Db 1168 TGAACCGAGGCGGCGTCTTGAATCATTTACTGGAGCGCTGAGAACCGCTCTCTTTCG 1227  
 Qy 158 aGlyValGlnGlyArgTyrProThrIleAspAla 169  
 Db 1228 TGGAGTTCAGGGTCGATACCCAGTATGTCGATGATGATGATGATGATGATGATGAT 1287  
 Qy 170 ----- 170  
 Db 1288 GACTCTCAGTACATTCCTCTCAATGTCCTACTATTACA-GACACCGAGATATCCTCTG 1346  
 Qy 179 aGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerAr 199  
 Db 1347 ACGGCAACCTGGGAGTCTTTCAGGGCTCTCTCAGTCTCTCTCTCTCTCTCTCTCTCT 1406  
 Qy 199 gValGlnSerLysAspPheSerLeuThrGluSerTyrGly-GlyHisLysTyrGly---- 217  
 Db 199 gValGlnSerLysAspPheSerLeuThrGluSerTyrGly-GlyHisLysTyrGly---- 217

Db 1407 AGTCAAGTCCAAAGAGTTCAACCTGTGGACAGAGAGTTACGGAGGTTGAGTCAACTTTC 1466  
 QY 217 ----- 217  
 Db 1467 ATACCAGACCGAGCTAAGCTGACTTGATCAAGACACTATGGACCGGAGTGTCTTT 1526  
 QY 218 -----ProAlaPhePheAsnHisPheTyrGluGlnA 228  
 Db 1527 TTCCTGGTGCACACATATGATCTAATGACCGAAGTCTTCAATCATTTCTACGAGCAAA 1586  
 QY 228 snGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyL 248  
 Db 1587 ATTCAAGATCGTAGCGGGAAGTCAATGGGCTCCAACTGAATTTTAACCTCCCTCGGA 1646  
 QY 248 leIleAsnGlyIleAspGluAlaIleGluAlaPro----- 260  
 Db 1647 TTATCAACGGCATCATGATGCGGCGGATTCAGGTACT-TAGAAATGCAGCTCCGCGAGAG 1705  
 QY 261 -----TyrTyrProGluPheA 266  
 Db 1706 GCTGCGCCTAGNAGGACATCGCTAAAGTAATTAATAGGACAGACTACTAGCGAGACTTTG 1765  
 QY 266 laValAsnAsnThrTyrGlyIleLysAla----- 275  
 Db 1766 CCGTTAATAATACATATGAATCAAAAGCTGTAAATTAATACACGATACATCGTGGATTT 1825  
 QY 276 -----ValAsnGluThrValTyrAsnTyrMetLysPhe 286  
 Db 1826 AGATCAACCGTCTCATGCTTCTAGGTCAATGACACAGCTGACACATATATGAAGTTC 1885  
 QY 287 AlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArg 306  
 Db 1886 GCCAACACGATGCCAATGATGCCAGGATCAGTTGCTCGTGTAAATTCACCAATAGG 1945  
 QY 307 ThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnVal 326  
 Db 1946 ACCTCGCTTCTGATTATGCTATATGTACAGAGACGCAATATATGTCAGGGACAATGT- 2004  
 QY 326 ----- 326  
 Db 2005 CGGTGAGTGTCTTCTGTCGAGGGGTGCAATGATGAAGAGACTTGTGAAGCTG 2064  
 QY 327 -----GluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHis 343  
 Db 2065 TCATGATACAGAGGCGCTTACTACAGTTTGGCGCGTGGCGGTATGATATCGGCAC 2124  
 QY 344 ProTyrAsp----- 346  
 Db 2125 CCCTACAA-TGTAAGTGGCAAGGATAAGGATTGTACTTCCGAACAGGACACTGCTCAT 2183  
 QY 347 -----AspProThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSer 362  
 Db 2184 ATGTCACGATAGGACCGACCGCCCGCTCTACTTTGTTGACTACCTCAAGAAAGACTCA 2243  
 QY 363 ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyr 382  
 Db 2244 GTCATGATGCTATCGCGTGGACATTAACATACACCGAGTCCAGCGGCGGAATATATAT 2303  
 QY 383 AlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIle 402  
 Db 2304 GCATTCCAGCAGACCGCGACTTTGATGGCGGAATTTCAATTGAGGACCTCGAGAGATC 2363  
 QY 403 LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp 422  
 Db 2364 CTCCAACTCCCGGTACGGTGTCTGTTGATCTACGGCGATGCCGACTATATCTGTAAGTGG 2423  
 QY 423 PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer 442  
 Db 2424 TTCGCGGTGAGGCGATCTCACTCGAGTTAACTACCCCATCGAGCTCAGTTCGGTGCA 2483  
 QY 443 AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 462  
 Db 2484 GCGGGATACACCCCATGACAGTAGATGGGGTTCGAATACGGGTGAGACTCGCGGAGTATGAC 2543

QY 463 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 482  
 Db 2544 AACTTTTCTGTCACCGCGGTATATCATGAGTGGCAGAGTTCCATACATATCAACGATC 2603  
 QY 483 AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTyrAspIleAlaGluGlyGlnLys 502  
 Db 2604 GCAGCGTTGCAGCTGTTCACCGTACTTTATTTGATGGATGATATTCGCGGGTACAACT 2663  
 QY 503 LysIleTyrProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer 522  
 Db 2664 CAGATTGGCCCGAATATAGCAACGAGGACATCGCAGGTACACACACGAGGTGCTTC 2723  
 QY 523 ValProLeuProThrAlaThrSer 530  
 Db 2724 GTGCCACTGTCCACGGGTGCGAGT 2747

RESULT 5  
 AX534872 1872 bp DNA linear PAT 22-NOV-2002  
 LOCUS Sequence 110 from Patent WO02068623.  
 DEFINITION AX534872  
 ACCESSION AX534872.1 GI:25261464  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Aspergillus niger  
 Aspergillus niger  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

REFERENCE  
 1 Edens, L., dijk Van, A. A., Krubasik, P., Albermann, K., Stock, A.,  
 Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W.,  
 Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S.,  
 Kemmer, W., Tan, P., Stiebler, J. and Albang, R.  
 Novel genes encoding novel proteolytic enzymes  
 Patent: WO 02068623-A 110 06-SEP-2002;  
 DSM N.V. (NL)  
 Location/Qualifiers  
 1. 1872  
 /organism="Aspergillus niger"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:5061"

TITLE  
 JOURNAL  
 FEATURES  
 source  
 1. 1872  
 /organism="Aspergillus niger"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:5061"

BASE COUNT 402 a 434 c 570 g 466 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6,68e-68 Length: 1872  
 Score: 980.50 Matches: 216  
 Percent Similarity: 56.23% Conservative: 91  
 Best Local Similarity: 39.56% Mismatches: 178  
 Query Match: 33.96% Indels: 61  
 DB: 6 Gaps: 19  
 US-09-712-338-2\_COPY\_19\_555 (1-537) x AX534872 (1-1872)  
 QY 26 ThrAlaAsnAsnValThrIleArgTyrLysGlu-----ProGlyAlaGlu----- 40  
 Db 94 ACGGCTGAGGATCTCAGCGTATTTCATTCGGAGATATTCCTCGGCGAGGATCTCCAT 153  
 QY 41 -----GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrVal 56  
 Db 154 AAGCAACCCCTCGGATCTGCACACACACCCCTCCACCCCGCTACTCCGGCTACATC 213  
 QY 57 AspThrSerProGluSer----- 62  
 Db 214 CACCTCCCGCCACACACCGCTTACCAATCTCTCCATCCAGGAATCAGATCTCCGAACA 273  
 QY 63 -----HisThrPhePheThrPheGluAlaArg-----HisAsnProGluThr 77  
 Db 274 TACCTATCAATACCTTTTTCGTGACTTCTCTCCGCCATCACCACAAATGATACA 333  
 QY 78 AlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPhe 97  
 Db 334 TCCCCACTACCATCTGGATGAACGCGCGCGCGCGGATCTCCATGATTTGGGCTATTT 393



QY	98	GlnduLeuGlyProCysHisValAsnSerThrPheAspAsp---TyrIleAsnProHis	111
DB	394	CAAGAAACGGGCCATGACTGTGAATACGGACATCGAATCCACGGCCCTATATCCCTGG	453
QY	117	SerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer	136
DB	454	TCGTGGAAATAGTACGTGCATATGTTGTATATGACGACGGTGCACACGGATTAGT	513
QY	137	TyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSer	156
DB	514	TATGATGTCITGAGCAATGGGACGTAGAT-----ITGAATCAGACAG	555
QY	157	PheAlaGlyValGlnGlyArgTyrProThrIleAsp-----AlaThrLeuIleAspThr	174
DB	556	TTT-----TTGGTGGGACGTCGCCAGTCAGGATGTCATGGGACGGTGAATGGGACG	609
QY	175	ThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGly-----	192
DB	610	GTTAAT---GGGGGAAGGGCGCTTTGGTGTGGTGGCAGGTTGGTGGGTGAATTCCT	666
QY	193	-----LeuProSerLeuAspSerArgValGlnSerLysAsp---PheSerLeuTrpThr	209
DB	667	GAATATGTTCTCTGTGTGCGGGAATGGTGGTGTATGACACGGGTGAGTATATGACG	726
QY	210	GluSerTyrGlyGlyHisTyrGlyProAlaPheAsnHisPheTyrGluGluAsnGlu	229
DB	727	CAGTCATATGGGGACGGTATGGACCGGCATACACGGCGCTCTTTCAGGAGATGAATGAG	785
QY	230	ArgIleAlaAsnGlySerValAsn---GlyValGlnLeuAsnPheAsnSerLeuGlyIle	248
DB	787	AGGATTCAGAGTGGGGAGGTAAAGCACCGGAAGAAGATCCATTTGGATACGCTGGCAT	846
QY	249	IleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAlaValAsn	268
DB	847	ATCAATGGGTGTGNGAATTTACTCTGCACAGTCCCTTCGTTCCTCAGCAGGGGTATAC	906
QY	269	AsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsn	288
DB	907	AATACGTATGGGTGAGGGAATCAATCGCACGCTCTACGACCGGGCTATGGATAGTTGG	966
QY	289	GlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAla	308
DB	967	AGCAAGCCTCGCGGGTGCAGGATATGATCATCGAGTGTCCGATGTCGCGAG-----	1020
QY	309	LeuAlaAsp---TyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGlu	327
DB	1021	CTCGAGATCCCTCATCTCGGAGGAGGCTCGGACACTGCTTCGCGGGAGATCAAG	1080
QY	328	GlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAsp	347
DB	1081	AGCCTGTATACGAATACCTCCGGCGAGGATACTACGACATAGCGCAITTCACGCCGAT	1140
QY	348	ProThrProSerSerTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIle	367
DB	1141	GCAGCTCTCGCTTACTTCGTCCGGTTCTTGAATCGCCCATGGGTGCAAAAGGCACTT	1200
QY	368	GlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPheGlnThr	387
DB	1201	GGSGTCCCGGTGAACATATACCATGTCTGACAGGCGAGTGGGAACACTTCGCTCCAGC	1260
QY	388	GlyAspPheValTrpPro-----AsnPheIleGluAspLeuGluIle	402
DB	1261	GGCGATTAT-----CCGCGAAATGATCCCGCGGAATGATCGGGGATATGCAIACCTG	1314
QY	403	LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp	422
DB	1315	CTTAGACTCCGGTGTCAAGTGGCTATGGTATATGGGACCGGACACTGCTTCTCGGTG	1374
QY	423	PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerClnAlaAlaGlnPheArgSer	442
DB	1375	CGCGGGGGGAAGATGTACCCGCTGCTGGGAGTACGAGGATCGGAGAAAGTTCCGTGCT	1434

QY	413	AlaGlyTyrThrProLeuLysValAsnGlyValGluTyr-----GlyGluThrArgGlu	460
Db	1435	GCTGGTATGCGAAGTCACACGAAGTCA---TCCACGTTGGGGTCTAGTAAGCGAG	1491
QY	461	TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln	480
Db	1492	TATGGGAACCTCTCGTCCACCGGTCTCTTCACGGCGGCCCAAGAGGTGCCATTTATCAG	1551
QY	481	ProfileAlaSerLeuGlnLeuPheAsnAArgThrIlePheGlyTyrAspIleAlaGluGly	500
Db	1552	CCCGAAACGGCGTATGAGATTTTAATCGCGCTCAGTTTAATTCGGATATGCGACGGGA	1611
QY	501	GlnLysLysIleTyr-----ProSerTyrLysThrAsnGlyThrAlaThrAlaHis	518
Db	1612	GGCATTTCTTCGGAGCAGAAATCAGAGCTATGGGACGGAGGACCGTCTGTAACGTGCAT	1671
QY	519	ThrGlnSerSerValPro	524
Db	1672	ATCAAAACGAAGTGCCG	1689
RESULT 6			
AX534815			
LOCUS	AX534815	Sequence 53 from Patent WO02068623.	3221 bp DNA linear PAT 22-NOV-2002
DEFINITION	AX534815		
ACCESSION	AX534815		
VERSION	AX534815.1	GI:25261346	
KEYWORDS			
SOURCE		Aspergillus niger	
ORGANISM		Aspergillus niger	
REFERENCE		Eukaryota: Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;	
AUTHORS		Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.	
		Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock A.,	
		Kimpel, E., Kluebauer, S., Wagner, C., Fritz, A., gusted Von, W.,	
		Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S.,	
		Kemmer, W., Tan, P., Stiebler, J. and Albang, R.	
TITLE		Novel genes encoding novel proteolytic enzymes	
JOURNAL		Patent: WO 02068623-A 53 06-SEP-2002;	
FEATURES		DSM N.V. (NL)	
source		Location/Qualifiers	
		1..3221	
		/organism="Aspergillus niger"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:5061"	
BASE COUNT	739 a	773 c	869 g 840 t
ORIGIN			
Alignment Scores:			
Pred. No.:	2,47e-67	Length:	3221
Score:	977.00	Matches:	213
Percent Similarity:	55.60%	Conservative:	90
Best Local Similarity:	39.08%	Mismatches:	172
Query Match:	33.84%	Indels:	70
DB:	6	Gaps:	17
US-09-712-338-2_COPY_19_555 (1-537) x AX534815 (1-3221)			
QY	41	GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro	60
Db	918	GGCATCTGCACCCACACCCCTCCACCCCGCAGCTACTCGGCTACATCCACCTCCGCCA	977
QY	61	Gluser-----His	63
Db	978	CACACCCCTTACCAATCTCCATTCAGGAATCAGCATCTCGCAACCATACCTTATCAAT	1037
QY	64	ThrPhePheThrPhePheGluAlaArg-----HisAsnProGluThrAlaProIleThr	81
Db	1038	ACCTTTTCTGTGTAATTCCTTCCCGCCATCACCACAAATGATACATCCCACTCACC	1097
QY	82	LeuTrpLeuAsnGlyClyProGlySerAspSerLeuIleGlyLeuPheGluGly	101
Db	1098	AUCTGGATCAAGGGGGCGGGCGGATCTCCATGATTGGCTATTTCAGAGAAGCGG	1157

QY 102 ProCysHisValAsnSerThrPheAspAsp---TyrIleAsnProHisSerTrpAsnGlu 120  
 DDb 1158 CCATGCTACTGTGAATACGACTCGAATCCACGGCTATATCCCTGGTCTGGTGAATGAG 1217  
 QY 121 ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 140  
 DDb 1218 TAGTCGATATGTTGATATTAGCAGCGGGTTCAGACGGGATTTAGTTAATGATGTTG 1277  
 QY 141 ValAspGlySerIleAsnProValThr----- 149  
 DDb 1278 AGCAATGGACGCTAGATTGGTTAGTGGGAGATAGATTAGTATTACTGAAGTGAG 1337  
 QY 150 -----GlyValValGluAsnSerPheAlaGlyValGlnGlyArgTyrProThrIle 167  
 DDb 1338 AGGATGGAGTAGCAGACAGATGAGACGTTT-----TTGGTGGGACGTTGGCAGTCA 1391  
 QY 168 Asp-----AlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIle 185  
 DDb 1392 GATGTGCATGGGACGCTGAATGGACGTTAAT---GGGGAAAGGGCGCTTTGGTTGG 1448  
 QY 186 LeuGlnGlyPheLeuSerGly-----LeuProSerLeuAspSerArgValGln 201  
 DDb 1449 TTGCAGGTTGGTGGTGGTGAATCTCTGAATATGTTCTCTCTGTCGCGGGAATGGTGGI 1508  
 QY 202 SerLysAsp---PheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 220  
 DDb 1509 GGTGATCAGACGGTGTGATATGACGGAGTCAATATGGGACGGTATGACCGGCATAC 1568  
 QY 221 PheAsnHisPheTyrGlnGluAsnGluArgIleAlaAsnGlySerValAsn---GlyVal 239  
 DDb 1569 ACGGCGCTCTTTCAGGAGATGAATGAGAGATTTCAGAGTGGGAGGTAAACCGGGAAG 1628  
 QY 240 GlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAla 259  
 DDb 1629 AAGATCATTTGGATCGCGGCATATTCATCGGGTGTGATTTACTCGTCGACGTC 1688  
 QY 260 ProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleIleAlaValAsnGluThr 279  
 DDb 1689 CCTTCGTTCCCTGACGAGCGGTATACAAATACGATATGGATCGAGGGAATCAATCGCAC 1748  
 QY 280 ValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSer 299  
 DDb 1749 CTCTACACCGCGGTATGGATAGTTGGAGCAAGCCIGCGGGTGTGAGGATATGATCATC 1808  
 QY 300 ThrCysLysGlnThr-----AsnArgThrAla 308  
 DDb 1809 GAGTGTCCGATGTCGCGAGCTCGGAGATCCCTCATGATATGCGGCAATGACACGGTA 1858  
 QY 309 LeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGly 328  
 DDb 1869 -----AATACATCTCGAGGAGCGCTCGGACTACTGTTCGCGGAGATCAAGAGC 1919  
 QY 329 ProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro 348  
 DDb 1920 CTGTATACGAATACCTCCGGCGAGGATACACATAGCATGCGCATTTACGCCGGATGCA 1979  
 QY 349 ThrProSerTyrTrpAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGly 368  
 DDb 1980 GCTCTCGTGTCTACTTCGTCGCGGTCTTGAATCGCCCATGCGGTGCAAAAGCATCTGG 2039  
 QY 369 ValAsnIleAsnTyrThrGlnSerAsnAspValTyrTyrAlaPheGlnGluThrGly 388  
 DDb 2040 GTCCCGGTGAATATACCATGCTGCTCAGAGCGAGTGGGGAACAGTTTCGCTCGACGGGC 2099  
 QY 389 AspPheValTrpPro-----AsnPheIleGluAspLeuGluGluIleLeu 403  
 DDb 2100 GATTAT-----CCGCGAATGATCCCGCGGATGATGTCGGGATATGGATCTGCTT 2153  
 QY 404 AlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPhe 423  
 DDb 2154 GACTCCGGTGTCAAGTGGCTATGATATGCGGACCGGACTATGCTGTGTCGGTGGCG 2213  
 QY 424 GlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAla 443

Db 2214 GCGCGGAGAGATCTACCCCTGCTGGTGGAGTACAGAGATCGGAGAGATTCCTGCTGCT 2273  
 QY 444 GlyTyrThrProLeuLysValAsnGlyValGluTyr-----GlyGluThrArgGluTyr 461  
 DDb 2274 GGTAATCCGAGAGTGGAGCAAGTCA---TCCTACGTTGGGGTCTAGTAGGCAAGTAT 2330  
 QY 462 GlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnPro 481  
 DDb 2331 GGCAACTCTCTGTTTCAGCGGTCTTTTCAGCGGGCCATGAGTGCCATTTATTCAGCC 2390  
 QY 482 IleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGln 501  
 DDb 2391 GAAACGCGATGAGATTTTAAATCGCGCTCAGTTAATGGGATATTCGACGGGAGGC 2450  
 QY 502 LysLysIleTrp-----ProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThr 519  
 DDb 2451 ATTTCCTGGAGCAGATCAGAGCTAATGGACGAGGACCGTCTGACGTGCAATATC 2510  
 QY 520 GlnSerSerValPro 524  
 DDb 2511 AAAACGAGTGGCG 2525

RESULT 7  
 AX534866  
 LOCUS  
 DEFINITION Sequence 104 from Patent WO02068623.  
 ACCESSION AX534866  
 VERSION AX534866.1 GI:25261452  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Aspergillus niger  
 Aspergillus niger  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

REFERENCE  
 1 Edens, L., dijk Van, A. A., Krubasik, P., Albermann, K., Stock, A.,  
 Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W.,  
 Heinrich, O., Maier, D., Spreafico, F., Folkers, O., Hopper, S.,  
 Kemmer, W., Tan, P., Stiebler, J. and Albang, R.  
 Novel genes encoding novel proteolytic enzymes  
 Patent: WO 02068623-A 104 06-SEP-2002;  
 DSM N.V. (NL)

TITLE  
 JOURNAL  
 FEATURES  
 Location/Qualifiers  
 1..1665  
 /organism="Aspergillus niger"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:5061"  
 374 a 455 c 424 g 412 t

BASE COUNT  
 ORIGIN

Alignment Scores:  
 Pred. No.: 7.48e-65 Length: 1665  
 Score: 941.00 Matches: 209  
 Percent Similarity: 51.75% Conservative: 72  
 Best Local Similarity: 38.49% Mismatches: 172  
 Query Match: 32.59% Indels: 90  
 DB: 6 Gaps: 13

US-09-712-338-2\_COPY\_19\_555 (1-537) x AX534866 (1-1665)

QY 13 GlnLeuProLysAsnProThrGlyValLysThrLeuThrAla-----AsnAsnVal 30  
 DDb 49 CAATTCCTCCGACGCGGAGGCAATCACTGTCTCAAGTCCAGTCCAGTGCATGAGNATG 108  
 QY 31 ThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyValLys 50  
 DDb 109 ACTATTTCTTTCAAGAGGCT-----GGAATTCGAAACTACGCGGTGTCCGA 159  
 QY 51 SerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePheTrpPheGlu 70  
 DDb 160 TCTTATTCGGGTATGTACACCTTCCCGCGCTCAACACGACTCTTTTGGTTTTCGAA 219  
 QY 71 AlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySer 90

```

Db      220  GCGCGAAGATCCCAACATGCGCTCGGCATCTGGCATATGCGGTGGGTC 279
QY      91   AspSerLeuIleGlyPheGluLeuGlyProCysHisValAsnSer----ThrPhe 109
Db      280  TCGTCGCTATGGGCTCTTGAAGATATAGGTCCTTGTCATTCATCAGACTCCAG 339
QY      110  AspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGln 129
Db      340  ACCACAGTCTCAATCTTGGAGTGGAAACATGAAGTCAATCTTATCTTCCCTTGACCAG 399
QY      130  ProLeuGlyValGlyPheSerTrpAspThrValAspGlySerIleAsnProValThr 149
Db      400  CCAACTCAAGTCGGGCTTCATACGATGTCACCAACAAATGGCACTTGACA----- 450
QY      150  GlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAla 169
Db      450  ----- 450
QY      170  ThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPhe 189
Db      451  -----GCTAATGGGACTGCATTCGCGCTCAGCTCATGCGATTCGCGCAAAACCTGG 504
QY      190  LeuSerGlyLeuProSerLeu-----AspSerArgValGlnSerLysAspPheSer 206
Db      505  TTTTTCGAGTTCCTCCACTACAGCAACACGATGATCGTGTG-----ACT 549
QY      207  LeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsnHisPheTyrGlu 226
Db      550  CFTCGGCTGAAAGTACGGAGCCATATGTCAGGATCTTCGGTTCCTTCCAAACAG 609
QY      227  GlnAsnGluArgIleAlaAsnGlySerVal---AsnGlyValGln---LeuAsnPheAsn 244
Db      610  CAGATGACAAAATCGCAGAGGGGACTGCAGAAGACGGTGCAGATATTGCTATCTGCAC 569
QY      245  SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTrpProGlu 264
Db      670  ACGCTTGCAATGTGACGGCTTGATGATATGGTGATCCCAAGAGAGCGCTTACATTACT 729
QY      265  PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMet 284
Db      730  TGGCCA-----TACAATAACGTA 747
QY      285  LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr 304
Db      748  AGGCTCGCCCT-----TCITCATTC 768
QY      305  AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp 324
Db      769  AACTCGGAGGCTTCGGGATCAGGCGCTCGCCTCGCAAGCGGCTTTGAAAGAACGGCAT 828
QY      325  -----AsnVal----- 326
Db      829  TCGGCTTGCCCTCACTCAGGAAGAATATCTCAATTCGGAGGCGCTTGCCTAGAA 888
QY      327  -----GluGlyPro-----TyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg 342
Db      889  TGGGAGATGCCCATCACCCTACTACACACCTTCAATCGCGGTGTGATGACATCGCC 948
QY      343  HisProTyrAspAspProThrProProSerTyrTrpAsnLysPheLeuAlaLysAspSer 362
Db      949  CATCTAAGAACGACCCATTCCTGCGCAAGCACATGCTCGGATATTTCACGAGGAGTCC 1008
QY      363  ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyr 382
Db      1009  GTCCTTCGGCTCTGGGTACCACTCAATTCATCGCTTCGAGTGGCGGTGGGTCACA 1068
QY      383  AlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIle 402
Db      1069  CAGTTCAATAAAACCTTTGATATGCGCCAGCGGCTTCCTGGATGCAATTTGGTACTCTC 1128
QY      403  LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp 422

```

```

Db      1129  CTCGACAGTGGTGTAAAGTACACATGATACGGAGATCGTGATATACGCTGCAATGG 1188
QY      423  PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer 442
Db      1189  GTCGGGGCGGAAAAAGCCAGCCITGCAGTTCCTGATTCCTGATTCCTGATTCCTGATTC 1248
QY      443  AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 462
Db      1249  ACGGATATCTCCCATCTTACGCCGCGGATCAGCGCATGACCCGACGCTGGGC 1308
QY      463  AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 482
Db      1309  AAGIACAGCTCTACTCGGTCTTCCAGCGGCGATGAGTCCCTCCCTACCGAGCTGTC 1368
QY      483  AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLys 502
Db      1369  GCGGCGTATGAGATCTTCATGCGGCGGACATTCACAAAGATATCCCTACTGCCCTTGG 1428
QY      503  LysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer 522
Db      1429  GCTGTGTGATGACGATTCAGTGGTGGACCTTAGGATACGTGGCATATCAAGAATATC 1488
QY      523  ValProLeu 525
Db      1489  CCTCCTATT 1497

RESULT 8
AX534824
LOCUS      1581 bp      DNA      Linear      PAT 22-NOV-2002
SEQUENCE 62 from Patent WO02068623.
ACCESSION AX534824
VERSION   AX534824.1  GI:25261366
KEYWORDS  .
SOURCE    Aspergillus niger
ORGANISM  Aspergillus niger
REFERENCE 1
AUTHORS   Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock, A.,
          Kimpel, E., Klugbauer, S., Wagner, C., Fritze, A., gustadt Von, W.,
          Heinrich, O., Maier, D., Spreafico, F., Folkers, J., Hopper, S.,
          Kemmer, W., Tan, P., Stiebler, J. and Albang, R.
          Novel genes encoding novel proteolytic enzymes
          Patent: WO 02068623-A 62 06-SEP-2002;
          DSM N.V. (NL)
FEATURES   Location/Qualifiers
            source
              1..1581
              /organism="Aspergillus niger"
              /mol_type="genomic DNA"
              /db_xref="taxon:5061"
BASE COUNT 343 a 473 c 423 g 342 t
ORIGIN

Alignment Scores:
Pred. No.:      5,986-63      Length:      1581
Score:          916.50      Matches:      212
Percent Similarity: 51.97%      Conservative: 78
Best Local Similarity: 37.99%      Mismatches: 187
Query Match:      31.75%      Indels:      81
DB:              6      Gaps:      13

US-09-712-338-2_COPY_19_555 (1-537) x AX534824 (1-1581)

QY      9   ValGlyArgArgGlnLeuProLysAsnProThrGlyValLysThrLeuThrAlaAsn 28
Db      46  GTCTCCCGGGCCCATTTGTGGCTCGGCCCGCATCTCATCCACCAAGGATATCTC 105
QY      29  AsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrProGly 48
Db      106  GACATCCCGCTCCCGCTACAAACAGATC---CCCACCGGCAITTTGTGAGACTGATCCCA 162
QY      49  ValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePheTrpPhe 68

```

```

Db      163  GTCAGAGCTTCCTCCGGTTAGCTCGATCTCGCTGAGCATGAGCACATCTTCCTTCCTGGTTC 222
QY
QY      69  PheGluAlaArgHis---AsnProGluThrAlaProIleThrLeuTrpLeuAsnGly--- 86
Db      223  TTCAGGGCGCAACCAAGATCCACCGAGGCTCCCTTGACCGTCTGGATCAATGAGGC 282
QY      87  -----GlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 102
Db      283  ATGCTGACCCCGGTCCTGGTTCCTCCCATGATCGGGTCTGTTCCAGAGACAGCCCA 342
QY      103  CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerIrrAsnGluValSer 122
Db      343  TCCGGCATGACGCCAATGGCTCGCTCTACAACACCCCTACTCTCGAACAACGCCAGC 402
QY      123  AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
Db      403  AACATGCTCATACATGACGAGCCGCTGACAGCCGCTCTCTCTACAGC----- 450
QY      143  GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
Db      451  -----ATTCCCGTCCCGGCTATGTGGATCTTCCACAGACAATGGTTTATGGGC 501
QY      163  ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaAla 182
Db      502  GCATTTCTCT----- 510
QY      183  TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
Db      511  -----CAGTACTCGCGC 522
QY      203  LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsn 222
Db      523  GAAACCTTCACCTTCACACGAGAGTATGCGGCCACTACGGCCCTCTTCACAGAG 582
QY      223  HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
Db      583  TACATCGAGGACGAGAACGCCCATCTCCAGCGGAGGC-----AAGAAGTCCAA 633
QY      243  PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyr 262
Db      634  CTGGCAGTGTGATGATCGCATGCGCATGGCTGTATGACCCGATATTCATACCGGCTAC 693
QY      263  ProGluPheAlaVal-----AsnAsnThrTyrGlyIleLysAlaValAsnGluThrVal 280
Db      694  TACAACCTTACGTATATCCGSCAACACATACAGTACCTGCTGCCATTCAACAGTCCATC 753
QY      281  TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr 300
Db      754  AGCTCGCTGATGTACAAACACCTCTATGGCCCGGAACTGCTCGACGCTCTACGAC 813
QY      301  CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn 320
Db      814  TGC-----GCCGCCGAGGATCGACGAGATCTGCAGACTGCCGAGAT 858
QY      321  MetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAsp 340
Db      859  TTTTGGCCCAACGAGGTGCAAAACGCTACGACATTTACTCCGCTCGGATGATGAC 918
QY      341  IleArgHisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys 360
Db      919  TTTTCGTGAATCACTACTCCGACCGCTCCCTACGAGTCTTACGTTGACTACCTCCGAACA 978
QY      361  AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal 380
Db      979  GCCTCGGTGAGCGCCCATCGGGCATACATCAATTACACGAGAGACAAACGCTGTT 1038
QY      381  TyrTyrAlaPheGlnGlnThrGly---AspPheValIrrProAsnPheIleGluAspLeu 399
Db      1039  GGACTCGCTTTTCTGCTCCACCGGTGACGAGCGGCGACTCATGACACCATCCAGAGTGT 1098
QY      400  GluGluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIle 419
Db      1099  GGCAAGTGTCTCAAAACAGGGTGTACCGGTGTCTATGTACCGCGGGGATGCCGACTATAC 1158

```

---

```

QY      420  CysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln 439
Db      1159  TCCAACCTGGCTGGTGGGAGCCGCTGCTGTTCAGGTC-----AAGGCCGCCAAC 1209
QY      440  PheArgSerAlaGlyTyrThrProLeuLysVal---AsnGlyValGluTyrGlyGluThr 458
Db      1210  TTCAGTAGTCGGGTACACCAACATGTCACTCGGATGGAGTGACACACCGCCAGG 1269
QY      459  ArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyr 478
Db      1270  CCCCAGGGGGCAATTTCCTTTGCGGAGTGTAGAGAGTGGACATGAGGTTCCTCTTC 1329
QY      479  TyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAla 498
Db      1330  TATCAACCTTCTGCTTCGCTGGAGATGTTTGGCGGCTCATTTGGCGGCAAGATGTGGG 1389
QY      499  GluGlyGlnLysIleIrrProSerTyrLysThrAsnGlyThr----- 513
Db      1390  ACGGAAAGATTCCCATCTCGTCGAGTTTACAGACGCTGGGACCGCCCAAGAGTTACTAC 1449
QY      514  -----AlaThrAlaThrHis---Thr 519
Db      1450  CGGGAGGCAACAGACAGATTCACTGAGGTGTTCGATCTCTGCGGAGTACACACA 1509
QY      520  GlnSerSerValProLeuProThrAlaThrSerMetSerSerValGlyMetAla 537
Db      1510  ACCACGATGCTCCGAACCCGGTGGAGCGGCTGAAGCGGATGGGACAGCT 1563

```

RESULT 9  
AX534767  
LOCUS  
AX534767  
DEFINITION  
AX534767  
ACCESSION  
AX534767.1 GI:25261241  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Aspergillus niger  
Aspergillus niger  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

REFERENCE  
1  
AUTHORS  
Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock, A.,  
Kimpel, E., Klubbauer, S., Wagner, C., Fritz, A., gustedt Von, W.,  
Heinrich, O., Maier, D., Spreafico, P., Folkers, U., Hopper, S.,  
Kemmer, W., Tan, P., Stiebler, J. and Albani, R.  
Novel genes encoding novel proteolytic enzymes  
Patent: WO 02068623-A 5 06-SEP-2002;  
DSM N.V. (NL)

FEATURES  
source  
1. .2940  
/organism="Aspergillus niger"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:5061"

BASE COUNT 691 a 818 c 732 g 699 t

ORIGIN

Alignment Scores:  
Pred. No.: 7.64e-61 Length: 2940  
Score: 894.00 Matches: 222  
Percent Similarity: 49.84% Conservative: 81  
Best Local Similarity: 36.51% Mismatches: 201  
Query Match: 30.97% Indels: 105  
DB: 6 Gaps: 16

US-09-712-338-2\_COPY\_19\_555 (1-537) x AX534767 (1-2940)

QY 9 ValGlyArgArgGlnLeuProLysAsnProThrGlyValLysThrLeuThrAlaAsn 28  
Db 498 GTCTCCCGGGCCCGAGTTTGGCTCCGCCACGAGTCTCATTCACCACGAGGATATCTC 557  
QY 29 AsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGly 48  
Db 558 GACATCCCGGCTCCGCTACAAACAGGTC---CCCACCGGCAATTTGAGACTGATCCAGT 614



Score:	887.50	Matches:	223
Percent Similarity:	48.15%	Conservative:	77
Best Local Similarity:	35.79%	Mismatches:	191
Query Match:	30.74%	Indels:	133
DB:	6	Gaps:	17
US-09-712-338-2_COPY_19_555 (1-537) x AX534809 (1-3080)			
QY	13	GlnLeuProLysAsnProThrGlyValLysThrLeuThrAla-----AsnAsnVal	30
DB	563	CAATTCCTCCCGACCGGAGGACATCACTGCTCAAGTCCAAAGTGCATGAGAAATGG	522
QY	31	ThrIleArgTyrLysGluProGlyAlaGlu-----	40
DB	623	ACTATTCTTCAAGACGTGTGCGAGA-GTATCTAGAAATAGCTTTTATGCTCGATGC	681
QY	41	-----GlyValCysGluThrThrProGlyValLysSerTyrSerGly	54
DB	682	CGTGTGATTGTGAGCTTGAATTCGGAAGTACGCGGGTGTCCGATCTTATTCGGC	741
QY	55	TyrValAspThrSerPro-----	60
DB	742	TATGTACACCTTCCCGCCGGTTCCTTCGACGGGACAGGACAGTGCAGGATTAATCT	801
QY	61	-----Glu-SerHisTh	64
DB	802	ATCAACACGTAAACCAATTCGAAACATTTGGAGGATGAGCAATTAAGCTCAACACG	861
QY	64	rPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLe	84
DB	862	CTCTCTTTGGTTTTCGAAGCCGCAAGATCCAGCAATGCGCCTTGSCCATCTGGCT	921
QY	84	uAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGlyProCysHi	104
DB	922	CAATGGCGGTCGGGGTGGCTGCTGCTATGGGGCTCCTTGGAAGAATTAGTCTCTGTC	981
QY	104	sValAsnSer---ThrPheAspTyrIleAsnProHisSerTrpAsnGluValSerAs	123
DB	982	CATTGCATCAGATCCCAAGACACAGCTCCTCAATCCTTGAGTTGGAACAATGAAGTCAA	1041
QY	123	nLeuLeuPheLeuSerGlnProGlyValGlyValGlyPheSerTyrSer---	138
DB	1042	TCTTCTATTCTTACACGCAACTCAAGTCGGCTTCTCATAGATGTCCCAACAATGG	1101
QY	139	-----AspThrValAspGlySerIleAsnProValThrGly-----	151
DB	1102	CACCTTGGTCGACTCGGACGGCGGAGAGAGATAGTTTCGGGTGATTTCCTCATGA	1161
QY	151	lValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLe	171
DB	1162	TGTTCCCGCAGTCCCAACTTCAACCATCATGTTGGTACCTTTGCAAGC-----	1215
QY	171	uIleAspThrThrAsn-----LeuAlaAlaGluAlaAlaTrpGluIleLeuGlnG1	188
DB	1216	TGCACAGACAGCTAATGGGACTGCATTCGGGGCTCAGCTCTATGGCATTTCCGCGCAAC	1275
QY	188	yPheLeuSerGlyLeuProSerLeu-----AspSerArgValGlnSerLysAspPh	205
DB	1276	CTGGTTTTCAGTTCCTCACACTACAGCCAAACGATCATGCTGC-----	1321
QY	205	eSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsnHisPheTyr	225
DB	1322	-AGTCTCTGGCTGAAAGTTACGGAGGCAATATGTTCCAGGCACTTTCGGTCTCTCCA	1380
QY	225	rGluGluAsnGluArgIleAlaAsnGlySerVal---AsnGlyValGln---LeuAsnPh	243
DB	1381	ACAGCAGAATCACAAAATTCGAGAGGGGACTTCGAGAGACGGTGCACAGTATTGCACT	1440
QY	243	eAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrPr	263
DB	1441	CGACAGCTTGGCATTTGACAGCGCTTGATGATATGTTGATCCCAAGAAGAGGCTTACAT	1500
QY	263	oGluPheAlaValAsnAsn-----	269

RESULT 11  
D86560  
LOCUS Schizosaccharomyces pombe cpy1 gene for carboxypeptidase Y,  
DEFINITION complete cds.  
ACCESSION D86560  
VERSION D86560.1 GI:3046860  
KEYWORDS cpy1; carboxypeptidase Y.  
SOURCE Schizosaccharomyces pombe (fission yeast)  
ORGANISM Schizosaccharomyces pombe

Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
Schizosaccharomycetales; Schizosaccharomycetaceae;  
Schizosaccharomyces.

REFERENCE  
AUTHORS

Tabuchi, M., Iwaihara, O., Ohtani, Y., Ohuchi, N., Sakurai, J., Morita, T., Iwahara, S. and Takegawa, K.  
Vacuolar protein sorting in fission yeast: cloning, biosynthesis, transport, and processing of carboxypeptidase Y from *Schizosaccharomyces pombe*  
J. Bacteriol. 179 (13), 4179-4189 (1997)

JOURNAL  
MEDLINE

97352672  
9209031  
2 (sites)  
Takegawa, K., Tabuchi, M., Iwaihara, O., Morita, T. and Iwahara, S.  
Cloning and characterization of cpyl gene from *Schizosaccharomyces*

## JOURNAL

3 (bases 1 to 4308)  
Takegawa, K.  
Direct Submission  
Submitted (16-JUL-1996) Kaoru Takegawa, Kagawa University,  
Department of Bioresource Science, Faculty of Agriculture;  
2393, Kita-gun, Miki-cho, Kagawa 761-07, Japan  
(E-mail: takegawa@ag.kagawa-u.ac.jp. Tel: 087-891-3116,  
Fax: 087-898-7295)

## COMMENT

on Apr 13, 1998 this sequence version replaced gi:2274843.  
Sequence updated (08-Apr-1998).

## FEATURES

```

1. .4308
/organism="Schizosaccharomyces pombe"
/mol_type="genomic DNA"
/db_xref="taxon:4896"
812. .3820
/gene="cpyl"
812. .3820
/gene="cpyl"
/codon_start=1
/product="carboxypeptidase Y"
/protein_id="BAA25568.1"
/db_xref="GI:3045861"

```

/qd\_xref="GI:304861" data-bbox="100 110 115 890">
EDP IREQESSAPLMNDEKDEAEAPKNGHNEFDDEDDVALEHPPKDKLDSF  
LQAPURLHTVSDRLNIFDDEDEHVRKEHRSADDEARRKKGKGGKHNKGG  
HAKGKKSHPKDEDSYFDDERAKNEFDDEDEERANNEGERMPPRMHNKPPG  
NMPKPMNNEHGEHMPRPMNNEHGEHMPRPMNNEHGEHMPRPMNNEHGEHMP  
MPPRPMNNEHGEHMPRPMNNEHGEHMPRPMNNEHGEHMPRPMNNEHGEHMP  
EMHPPRMKNEHEHGEHMPRPMNNEHGEHMPRPMNNEHGEHMPRPMNNEHGEH  
KSPVDEHNNKPKDKENHKGKDEHNNKPKDKENHKGKDEHNNKPKDKENHKGKDEH  
NKPVEDEAEFSRFDLTGTYGNTVHIOENAKPDSRYAFDFDDELLDLP  
EOPATYLEMKACGIDMTAFRDSHPKSNAPQADSRYPVQSENGEHNKAF  
POHTLKYKSPESLGDITVKQTYGLDVEDDRHLFEFDEPNDEPVLVLGG  
PCCSLIIGFWELPESINLETIKPEYNHSWNSAVFLDQINTGSPQDVSLLD  
TVAGKDVATFLNLEFAKPPHAYDLHFIAGESYAGHYIQPAKEIMHNGGANFVA  
SYEMEKYGLKSVLIGNLDTPLVYGYKMACSPGYTMSQERDGTGAYDT  
CAKLTJCYGTGTGTPWCIGLSCYNNAMIGFTKILNLYIREECRDEHLCYPTG  
ALSYLNOGEVQALGYEYDYGKCTEIVNTGFTGKGDWMTFRDDYVATILEAGPLV  
LYAGDADYCNYNAGNEAWTDLNAGWOREYAEALKPWSNPGKAGGCKSKFNQGYLR  
LYEAGHWPNPEASLEMLNWSNDSIFA"

[illegible]

	Alignment Scores:		
aligned No.:	1 31e-41	Length:	4308
% Covered:	652.50	Matches:	170
Percent Similarity:	48.47%	Conservative:	67
Best Local Similarity:	34.76%	Mismatches:	168
Query Match:	22.60%	Indels:	85
DB:	8	Gaps:	15

US-09-712-338-2\_COPY\_19\_555 (1-537) x D86560 (1-4308)

31 ThrileArgtyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyVallys 50

Db	2519	ACTCTTCGCGTTAAAGATTCCAAACCTGAATCACTTGTTGATTTGACACT-----GTTAAG	2579
Qy	51	SerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePheTrpPhePheGlu	70
Db	2573	CAATACACCGGTTACTTATAGATGTCGAGATGACAGACATCTTTCTTCTGGTGTCTTTGAA	2632
Qy	71	AlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySer	90
Db	2633	TCTAGAATGATCCCGAGATGATCCCGICGTGTTGGTIGAACGGGTGCTCGTGTGC	2692
Qy	91	AspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsn---SerThrPhe	109
Db	2693	TCCTCCCTTACTGCTTTGTTTCANGAATATAGTCTCTTCTCAATCAACATTTGAGCAA	2752
Qy	110	AspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGln	129
Db	2753	AAACCCGAATATACCCICACAGTTGGAATCCAAATGCTTCAGTATATCTTTTGGATCAA	2812
Qy	130	ProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnProValThr	149
Db	2813	CCATCAACACGGGTTTCAGCAACGGAGAT-----GACTCGGTCTTGACACTGTTACG	2866
Qy	150	GlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAla	159
Db	2866	-----2866	
Qy	170	ThrLeuIleAspThrTrpAsnLeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPhe	189
Db	2867	-----GCVGTTAAGAGATGTTTATGC-AFTCTTGAACCTTTT	2901
Qy	190	LeuSerGlyLeuProSerLeuAspSerArgValGlnSer-LysAspPheSerLeuTrpTh	209
Db	2902	CITGCGCAAGTTCCTCA-----GTACGCTCATTGCGACTTTGCACATTCCTCG	2949
Qy	209	rgLusertGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGluGlnAsnG	229
Db	2950	TGAATCCTATGCTGGCCATATACATCCCCAGTTTGCCACAGAAATATGACGATCAACCA	3009
Qy	229	uArg-----IleAlaAsnGly---SerValAsnGlyValGlnLeuAsnPheAs	244
Db	3010	AGCTGCTAATCTTTTGTGCGCAGCGCTATGAAATGGACAGCAATATCAATTTGAA	3069
Qy	244	nSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProG	264
Db	3070	GAGTGCTTCATGTGAATGTTTGACTGATCCTTTGGTCCATACTACTTTTACGAAA	3129
Qy	264	uPheAlaValAsnAsnThrTyrGly-----IleuY	274
Db	3130	ATGGCTTCGGAGCGCTTACGGTCTTATATGCCCCAGAGGAATGTGATCGCATATC	3189
Qy	274	sAlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCy	294
Db	3190	TGTTGCTATGATACC-----TG	3207
Qy	294	sGlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLe	314
Db	3208	CGTAGCTAATCACTGGCGTGTACCAGACTGGCTTACT-----CCGT	3252
Qy	314	uCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAl	334
Db	3253	TTGCATTTGGTGCTCTTTGTATGCAATAACGCTATGATTTGGACCA---TTTACTAAGAC	3309
Qy	334	agLysArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyr--	353
Db	3310	IGSACTCAACATTATGATATTCGTGAGAAATGCGGTGACCAAGAGCACTATGCTAACC	3369
Qy	354	-----TyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyVa	369
Db	3370	CGAAACCGGTGCAATTGAGTTACTTGAACCAAGAAITGTTCAAGAAGCTTTGGGAGT	3429
Qy	369	IAsnIleAsnIlyThrGlnSerAsnAspValTyrTyrAlaPheGlnGlnThrGlyAs	389
Db	3430	TGAATACCATACAGGATCCATPACTGAAGTAAACATTCGTTTCCCTTTCAAGGTTGA	3489





```

/misc_feature
/feature="SPAC19G12.02c"
/note="gtatgt, splice donor sequence"
complement(4240..4253)
/misc_feature
/feature="SPAC19G12.02c"
/note="ctaacgaatgtag, splice branch and acceptor"
complement(4336..4341)
/misc_feature
/feature="SPAC19G12.02c"
/note="gtatgt, splice donor sequence"
complement(4468..4488)
/misc_feature
/feature="SPAC19G12.02c"
/note="PS00058 DNA mismatch repair proteins mutL"
5410..6372
/misc_feature
/feature="SPAC19G12.03"
5410..6372
/misc_feature
/feature="SPAC19G12.03"
len:320, LOW SIMILARITY:Yersinia pseudotuberculosis, CAB46604, hypothetical 33.7 kD protein, (293 aa), fasta scores: opt: 155, E():0.003, (27.5% identity in 265 aa)
/misc_feature
/feature="SPAC19G12.03"
/codon_start=1
/label="SPAC19G12.03"
/product="possible polysaccharide deacetylase"
/protein_id="CAB10114.1"
/db_xref="GI:2239196"
/translation="MYETRLTGNAGKPVDTNPNNKSIKIAVSEFVNVYEGGERSLLY
EDGFTFTLACGLMPPNRPVRSRSTESCFYSGSCGFWRLINLKKHKVPTCNAL
GOAEKNPVVVGAMEAGCEVGSISHRWINEGVPPTEYEHKKSVOALOKASPSNS
APRSWTGRASLNRKLVCOVYKDLGPQFDSDEYNDVYVWADPLASKPGAEDDK
GLLIVPTLVNDMKYAVAPFCNSDDFTYARDADFLVYEEGLEGAPKMMTGLHCR
LITGRGRFGLQKLMHEITSKEGVVATREQLAQWASAKHPYKA"
5608..6042
/misc_feature
/feature="SPAC19G12.03"
/note="Pfam match to entry PF01522 Polysac_deacet"
6888..7463
/misc_feature
/feature="SPAC19G12.04"
6888..7463
/misc_feature
/feature="SPAC19G12.04"
len:191, SIMILARITY:Saccharomyces cerevisiae, DAL3_YEAST, ureidoglycolate hydrolase, (195 aa), fasta scores: opt: 360, E():5e-17, (36.5% identity in 200 aa)
/misc_feature
/feature="SPAC19G12.04"
/codon_start=1
/label="SPAC19G12.04"
/product="putative ureidoglycolate hydrolase"
/protein_id="CAB10115.1"
/db_xref="GI:2239197"
/db_xref="SPTREMBL:O13843"
/translation="MEAKKIYAQALIDEAFAPFGSVVQKDDVKMVSANGSIKKYKLV
VSEIQNYEKSSASTRGVWFFSTHPSVHPANDEHAAFOISLVRHPFTTQTFPM
CRSDQAYLIAVAPNPDMPDWNQTAQVAKGAQGVITYSAGVWHAPWYTGKTMVL
AAFNVENGVAEEDCOVQSTESPIEFIKIST"
8397..9272
/misc_feature
/feature="SPAC19G12.05"
8397..9272
/misc_feature
/feature="SPAC19G12.05"
len:291, SIMILARITY:Saccharomyces cerevisiae, TXP_YEAST, tricarboxylate transport protein, (299 aa), fasta scores: opt: 738, E():0, (49.0% identity in 290 aa)
/misc_feature
/feature="SPAC19G12.05"
/codon_start=1
/label="SPAC19G12.05"
/product="putative tricarboxylate transport protein"
/protein_id="CAB10116.1"
/db_xref="GI:2239198"
/db_xref="SPTREMBL:O13844"
/translation="MSTVAKINPKSNKPGPVKSIAGGVAGAEIISITYPAEFAKIR
LQLYRVEGTAKLPFGLEWYGVNGLSKAAVRFAPDSFKKSLSDHGHILT
GPRVLGAGVAESVLVLTPTESIKTAI IDDRKPNRLKGFQASRIYVHENGIR
GLYRLGAATVAQOANSVGRFTAYNSIKOSLQSLPPDEKLSITVITFLVSGVAGIIV
YCTQPTDITVKSQMSLSASKEYKNSIHCAYKILITQDLRFWSGATPRLARLILSGGI
VTVTEKVMELIKPF"
8397..9020
/misc_feature
/feature="SPAC19G12.05"
/note="Pfam match to entry mito_carr PF00153, Mitochondrial carrier proteins"
8505..8531
/misc_feature
/feature="SPAC19G12.05"
/note="PS00215 Mitochondrial energy transfer proteins signature"
complement(10121..10516)

```

Db	19409	TGGTGCTATGATACC-----TG	19392
Qy	294	sGlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLe	314
Db	19391	CGCTAAGCTAACTACCTGGCTGTACAGACTGCTTACT-----CCTGT	19347
Qy	314	uCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAl	334
Db	19346	TTGCATTGGCTCTTTGATATGCAATAACGCTATGATGGACCA--TTTACTAAGAC	19290
Qy	334	aglyArgGlyValTyrAspIleArgHisProTyrAspAspProThrProSerTyr--	353
Db	19289	TGGACTCAACATTTATGATATTCCTGCAAGAAATGCCGTGACCAAGAGCATTATGCTACCC	19230
Qy	354	-----TyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyVa	369
Db	19229	CGAAACGGTGGCAATTGAGAGTTACTTGAAACCAAGAAATTTGTCAAGAAGCTTTGGAGT	19170
Qy	369	lAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPheGlnGlnThrGlyAs	389
Db	19169	TGAATACGATTACAGGGATGCAACTACTGAAGTAAACATTGGTTCTCTTTCAAGGGTGA	19110
Qy	389	pPheValTrpProAsnPheIleGluAspLeuGluGluIleLeu-----AlaLeuProVa	407
Db	19109	CTGGATCGGTAAGACTTCCGTGACGATGTCACCGCAATCTTAGAAGCTGGCCTACCCGT	19050
Qy	407	lArgValSerLeuIleTyr---GlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyG1	426
Db	19049	T-----CTTATCTATCCCGTGTGCTGACTACATTGCAATTACATGGCAATGA	18999
Qy	426	nAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGlyTyrTh	446
Db	18998	AGCTTGGACCGCGCACTTACGTGGCTGCTGCAACGTGACTTTTATGAGCGCAATGAA	18939
Qy	446	rProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPh	466
Db	18938	GCCTTGGAGTCCCTAATGGAAGGAAGCTGCTGCTGTAAGTCTTTCAAAAACACTTGTGA	18879
Qy	466	eThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuG1	486
Db	18878	TCYTGGCTCTACGAAGCTGGACACATGCTTCTCTTCAACCAACCGCAAGCTATTAGA	18819
Qy	486	nLeuPheAsnArgThrIlePheGly 494	
Db	18818	AATGTTGAACAGCTGGATAGATGTT 18794	
RESULT 13			
AX594884		1527 bp	DNA
LOCUS			linear
DEFINITION	Sequence 538 from Patent EP1258494.		PAT 14-FEB-2003
ACCESSION	AX594884		
VERSION	AX594884.1	GI:28396526	
KEYWORDS	Saccharomyces cerevisiae (baker's yeast)		
SOURCE	Saccharomyces cerevisiae		
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
REFERENCE	Saccharomycetales; Saccharomycetaceae; Saccharomyces.		
AUTHORS	Bauer, A., Gavin, A.C., Grandi, P., Krause, R., Kruse, U., Kuester, B.,		
TITLE	Marzioch, M., Schultz, J. and Superti-Furga, G.		
JOURNAL	Multiprotein complexes from eukaryotes		
CELLZONE	Patent: EP 1258494-A 538 20-NOV-2002;		
AG (DE)	CELLZONE AG (DE)		
FEATURES	Location/Qualifiers		
source	1. 1527		
	/organism="Saccharomyces cerevisiae"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:4932"		
BASE COUNT	439 a 300 c 319 g 469 t		
ORIGIN			
Alignment Scores:	7.21e-41	Length:	1527
Pred. No.:			

Score:	636.00	Matches:	167
Percent Similarity:	46.29%	Conservative:	70
Best Local Similarity:	32.62%	Mismatches:	197
Query Match:	22.03%	Indels:	78
DB:	6	Gaps:	14
US-09-712-338-2_COPY_19_555 (1-537) x AX594884 (1-1527)			
Qy	1	LeuProGlySerThrProAlaSerValGlyArgGlnLeu-----ProLys	16
Db	127	CTTCCACAGATAACACAGCAACCCCTTAATGGAACGTTTGAATCAGCATGATCCGCTG	186
Qy	17	AsnProThrGlyValLysThrLeuThrThrAlaAsnAsnValThrIleArgTyrLysGlu	36
Db	187	TTTACAACATTTTATTTCTTCTGTCGACAGATTACAGTTTGAGACTTAGAAGTAGATAGAT	246
Qy	37	ProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrVal	56
Db	247	CCCTTCTAACTAGGAAT-----GACACCGTAAACAAATGTCGGGTACATG	294
Qy	57	AspThrSerProGluSerHisThrPhePheTrpPhePheGluAlaArgHisAsnProGlu	76
Db	295	GACTATAAGGATTCACCAACACTTTTTTACTGGTTTTTGAAGTAGAAGATCCTGCT	354
Qy	77	ThrAlaProIleThrLeuThrLeuAsnGlyClyProGlySerAspSerLeuIleGlyLeu	96
Db	355	AACGACCAATATATCTTTGGTTAAATGGTGGACCTGCTTCTCTGCTTACTGGGTG	414
Qy	97	PheGluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHis	116
Db	415	CTATTGCACTAGGCCCTCATCAATTTGGCCCGCATATGAACCAATCCCAATCCCTAT	474
Qy	117	SerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer	136
Db	475	TCITTGAATAATAACGCTTCAATGATCTTCTTAGAAGCCACTCGGAGTCGGCTTTTCC	534
Qy	137	TyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSer	156
Db	535	TATGGTGTAT-----	543
Qy	157	PheAlaGlyValGlnGlyArgTyr-ProThrIleAspAlaThrLeuIleAspThrThrAsn	176
Db	544	-----GAAAAGTCTCTCTACAAAA	564
Qy	177	LeuAlaAlaGluAlaAlaTrpGluLeuGlnGlyPheLeuSerGlyLeuProSerLeu	196
Db	565	TTAGCAGGCAAGATGCGTACATTTCTCTGGAATTTTGTGAAAGCTTTTCCCTCATTTA	624
Qy	197	AspSerArgValGlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyr	216
Db	625	-----CGCTCCACAGATTTCCACATTCGACGCGAAATCCTATCGAGGACATAT	672
Qy	217	GlyProAlaPhePheAsnHisPheTyrGluGlnAsn---GluArgIleAlaAsnGlySer	235
Db	673	ATCCCTCAATTCACATGAGATCGTTGTCAAGAACCTGAAAGA-----	717
Qy	236	ValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGlu	255
Db	718	-----ACGTTCATTTAACTTCACTTAATGATGTTGATTAATGATGATCATCAGACCT	765
Qy	256	AlaIleGlnAlaProTyrTyrProGluPheAlaVal---AsnAsnThrTyrGlyIleLys	274
Db	766	TGATTTCAAGCAGATTATTAYCAACCAATGCGATCGGGAAGGGGGCTATCACCTGTT	825
Qy	275	AlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCys	294
Db	826	CTCTCATCAGAGAAGATGTGAGAAAAATGAGTAAGCTGCGAGGTCGT-----	870
Qy	295	GlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu	314
Db	871	-----TGCTAGGTGAACAAAGTTATGTTATGCTTCTTAATCAAGT	912
Qy	315	-----CysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAla	332

```

Db 913 TTACCAATGATAGTCGCGCACTGCTTACGTGACTCTGCACCTTTGGAAACCGTACATTAAC 972
Qy 333 PheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro----- 348
Db 973 ---ACAGACTCAACGCTATGACATTAAGAGGGCCCTGTGAAGATTAATAGTACTGATG 1029
Qy 349 -----ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAsp 365
Db 1030 ATGTGTTATACAGTTCGCGTATGTCGACCAAGTATGAATTTCTGAAAGTTCAGAA 1089
Qy 366 AlaIleGlyValAsnIle---AsnTyrThrGlnSerAsnAsnAspValTyrAlaPhe 384
Db 1090 ACGTAGGTCGCGAGCTGATTAATATATCTGGCTGTGATAATGAGTGTTCACCGGAT 1149
Qy 385 GlnGlnThrGlyAspPheValTyrProAsnPhelIleGluAspLeuGluGluLeuAla 404
Db 1150 TTGTTTACGGCGGATGAAGTAACCA---TTTCAACAATATATGCTGTAATTTATAA 1206
Qy 405 LeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTyrPheGly 424
Db 1207 CACAACATTCGGTATTAATATATATGCGGTGATGAAGTATATATTTGTAATGGCTGG 1266
Qy 425 GlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaLalaGlnPheArgSerAlaGly 444
Db 1267 AACCATGCTTGGTCCATGACTTGGAAATGGAATCAATAAACGTAGTATCAGAGAAGG 1326
Qy 445 TyrThrPro-----LeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 462
Db 1327 TTAAGACCATGGGTCAGTAAGAACAGAGGTGGACAAAGTCAAGAACTATGGC 1386
Qy 463 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 482
Db 1387 CCTTTCACCTTTTGGAAATATACGATCCGGTATATGTCGCTATGATGATCAACCG 1446
Qy 483 AlaSerLeuGlnLeuPheAsnArgThrIlePheGly 494
Db 1447 GCAAGTTTGGAAATGGTCAACAGTGGATTTCCGGT 1482

RESULT 14
SCYBR139W 2016 bp DNA linear PLN 11-AUG-1997
LOCUS S.cerevisiae chromosome II reading frame ORF YBR139W.
ACCESSION Z36008 Y13134
VERSION Z36008.1 GI:536435
KEYWORDS Saccharomyces cerevisiae (baker's yeast)
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
REFERENCE 1 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
AUTHORS Feldmann,H., Aigle,M., Aljinovic,G., Andre,B., Baclet,M.C.,
Barthe,C., Baur,A., Becam,A.M., Biteau,N., Boles,E., Brandt,T.,
Brendel,M., Bruckner,M., Bussereau,F., Christiansen,C.,
Contreras,R., Crouzet,M., Czapluch,C., Demolis,N., Delaveau,T.,
Daignon,F., Dondy,H., Duesterhus,S., Dubois,E., Dujon,B., El
Bakkoury,M., Entian,K.D., Feuermann,M., Fiers,W., Foho,G.M.,
Fritz,C., Gassehuber,H., Glansdorff,N., Goffeau,A., Grivell,L.A.,
de Haan,M., Hein,C., Herbert,C.J., Hollenberg,C.P., Holmstrom,K.,
Jacq,C., Jaquet,M., Jauniaux,J.C., Joniaux,J.L., Kalliesoe,T.,
Kiesau,P., Kirchuth,L., Koetter,P., Korol,S., Liebi,S., Logghe,M.,
Lohan,A.J.E., Louis,E.J., Li,Z.Y., Maat,M.J., Mallet,L.,
Mannhaupt,G., Messenguy,F., Miosga,T., Molemans,F., Mueller,S.,
Nasr,F., Obermaier,B., Perea,J., Pierard,A., Piravandi,E.,
Pohl,F.M., Pohl,I.M., Potier,S., Profit,M., Purnelle,B., Ramezani
Rad,M., Rieger,M., Rose,M., Schaaff-Gerstenschlaeger,I.,
Schrens,B., Schwarzlose,C., Skala,J., Slonimski,P.P.,
Smits,P.H.M., Souciet,J.L., Steensma,H.Y., Stucka,R.,
Urrestarazu,A., van der Aart,Q.J., van Dyck,L., Vassarotti,A.,
Vetter,I., Vierendeels,F., Vissers,S., Wagner,G., de Wergifosse,P.,
Wolfe,K.H., Zagulski,M., Zimmermann,F.K., Mewes,H.W. and Kleine,K.
Complete DNA sequence of yeast chromosome II
EMBO J. 13 (24), 5795-5809 (1994)

```

```

MEDLINE 95112788
PUBMED 7813418
REFERENCE 2 (bases 1 to 2016)
AUTHORS Becam,A.M., Herbert,C.J., Nasr,F., Slonimski,P.P. and Zagulski,M.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 2016)
AUTHORS MIPS
TITLE Direct Submission
JOURNAL Submitted (30-AUG-1994) Data collected by MIPS on behalf of the
European yeast chromosome II sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
PEAIDRES Location/Qualifiers
source 1..2016
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/chromosome="II"
328..1854
/note="ORF YBR139W"
/codon_start=1
/protein_id="CAA85097.1"
/db_xref="GI:536436"
/db_xref="SWISS-PROT:P38109"
/translation="MKYLNVFVLQLLISIKYASFGRAFSLEDDTFANLDKQLKP
QNTQOTLKLRLNDDPLFTTFSVDYDILRLKRVDPKSLGIDTVKVGMDYKD
SKHEFFYFFESRNDPANDPIILWNGSGPCSFTGLLFELGESSIGADMKPIHNPYSW
NNASMTFLQPLGVGFSGYDEKVSSTKLAKDAYIFLELFEAFPHLRNDFHAGE
SYAGHYTPQIAHEIVVKNPRTFNLTSMVINGITDPLIQADYVEPMACGGYHPVL
SSECEKMSKAAGRCRLKLCYKSLPCIVATYCDLSALLEPYINTGIVNDIRG
PCDNDGMCYTCGLRYVDVDMNEPEVQETLGSVDHNSGCDNDVFTGFLTGDSRP
FQOYIAELNHNIPVLIVYAGDKYICNWLGNHWSNELEWINKRRIQRMRLRPWWSKE
TGEELGVKNYGPFTFLRIYDAGHMVYDQPEASLEMVNSISGNRAFSDLSTLENAS"
BASE COUNT 608 a 379 c 392 g 637 t
ORIGIN
Alignment Scores:
Pred. No.: 1,02e-40 Length: 2016
Score: 636.00 Matches: 167
Percent Similarity: 46.29% Conservative: 70
Best Local Similarity: 32.62% Mismatches: 197
Query Match: 22.03% Indels: 78
DB: 8 Gaps: 14
US-09-712-338-2_COPY_19_555 (1-537) x SCYBR139W (1-2016)
Qy 1 LeuProGlySerThrProAlaSerValGlyArgGlnLeu-----ProLys 16
Db 454 CTCCACAGATACACAGCAACCCCTTAAATGGACCGCTTTGAATCAGCATGATCCGCTG 513
Qy 17 AsnProThrGlyValLysThrLeuThrAlaAsnAsnValThrIleArgTyrLysGlu 36
Db 514 TTTACAACITTTATTTCTCTGTGGACACAGATTACAGTTTGAGACITTAGACAGTAGAT 573
Qy 37 ProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrVal 56
Db 574 CCTTCTAAACTAGGAATT-----GACACCGCTAAACAAATGCGGTGATACATG 621
Qy 57 AspThrSerProGluSerHisThrPhePheThrPheGluAlaArgHisAsnProGlu 76
Db 622 GACTATAGAGTATCCCAACACTTTTTTACGGTTTTTTGAAGTAGAGGATTCCTGCT 681
Qy 77 ThrAlaProLleThrLeuThrLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeu 96
Db 682 AACGACCCCAATATTCTTTGTTAAATGSGTGGACCTGTTCTTCTCTGTTACTGGTGTG 741
Qy 97 PheGluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHis 116
Db 742 CTATTTGAACTAGCCCTCATCAATTTGGCGCATATGAACCAATCCCAATCCCTAT 801
Qy 117 SerThrAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer 136

```

Db 802 TCTTGAATAATACGCTTCAATGATCTTCTTAGAACAGCCACTCGGAGTCGGCTTTC 861  
Qy 137 TyrSerAspThrValAspGlySerIleAsnProValThrGlyValGluAsnSerSer 156  
Db 862 TATGGTGAT----- 870  
Qy 157 PheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrAsn 176  
Db 871 -----GAAAAAGCTCTCTACAAAA 891  
Qy 177 LeuAlaGluAlaAlaTrpGluLeuGlnGlyPheLeuSerGlyLeuProSerLeu 196  
Db 892 TTACAGCAAGATGGTACATTTCTCGAATGTTTTTGAAGCTTTTCTCATTTA 951  
Qy 197 AspSerArgValGlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyHisTyr 216  
Db 952 -----CCTCCACAGATTCCACATTCGACGGCAUCCCTAAGCAGACATAT 999  
Qy 217 GlyProAlaPhePheAsnHisPheTyrGluGlnAsn---GluArgIleAlaAsnGlySer 235  
Db 1000 ATCCCTCAATTCACATGAGATCGTTGTCAAGAACCTGAAAGA----- 1044  
Qy 236 ValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGlu 255  
Db 1045 -----ACGTTCAATTTAACTTCAGTTATGATGTTAATGGTATACACAGACCT 1092  
Qy 256 AlaIleGlnAlaProTyrTyrProGluPheAlaVal---AsnAsnThrTyrGlyIleLys 274  
Db 1093 TTGATTCACGACGATTTATTAAGAACCAATGCGATCGGGAAAGGGGCTATCACCCTGTT 1152  
Qy 275 AlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCys 294  
Db 1153 CTCCTATCAGAAAGATGTGAGAAAATGAGTAAAGCTGCAGGTCTG----- 1197  
Qy 295 GluAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu 314  
Db 1198 -----TCTCGTAGTTGACCAAGTATGTTATGCTTCTCAATCTAAGT 1239  
Qy 315 -----CysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAla 332  
Db 1240 TTACATGCTAGTCGCGACTGCTTACTGTGACTCTGCACCTTTGGACCGTACATTAAC 1299  
Qy 333 PheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspPPro----- 348  
Db 1300 ---ACAGACTCAACGCTATGACATTAGAGGGCCCTGTGAAGATAATAGTACTGATGCT 1356  
Qy 349 -----ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAsp 365  
Db 1357 ATGTGTTATACAGTCTCCGCTATGTCACGAGTATATGAATTTTCTGAAAGTCAAGAA 1416  
Qy 366 AlaIleGlyValAsnIle---AsnTyrThrGlnSerAsnAspValTyrTyrAlaPhe 384  
Db 1417 ACGTAGGTCGAGCGTCATTAATTTCTGGCTGTGATATGACGTGTTCCACCGATTT 1476  
Qy 385 GlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluGluIleLeuAla 404  
Db 1477 TTGTTTACGGCGCATGGAAGTAACCA---TTTCAACAATATATGCTGAATTAAT 1533  
Qy 405 LeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTyrPheGly 424  
Db 1534 CACAAATTCGGTATTAATATATATGCGGTGATAGGATATATTTGTAATGCTGGGA 1593  
Qy 425 GlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGly 444  
Db 1594 AACCATGCTGTGCTCAATGAGTTGGATGGATCAATAAACCCTAGGTATCAGAGAAGATG 1653  
Qy 445 TyrThrPro-----LeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 462  
Db 1654 TTAAGACCATGGGTGAGTAAGAAACACAGGTGAAGAGTTGGCAAGTCAAGAACTATGGC 1713  
Qy 463 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 482  
Db 1714 CCTTTCACCTTTTGAAGATATACGATGCGGTATATGCTGCTATGATCAACACCGAG 1773

Qy 483 AlaSerLeuGlnLeuPheAsnArgThrIlePheGly 494  
Db 1774 GCAAGTTTGGAAATGGTCAACAGTTGGATTTCGGGT 1809  
RESULT 15  
AX536454 2027 bp DNA linear PAT 22-NOV-2002  
LOCUS Sequence 55 from Patent WO02064766.  
DEFINITION  
ACCESSION AX536454  
VERSION AX536454.1 GI:25262853  
KEYWORDS  
SOURCE  
ORGANISM  
Saccharomyces cerevisiae (baker's yeast)  
Saccharomyces cerevisiae  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
REFERENCE  
AUTHORS Contreras,R.H., Eberhardt,I., Luyten,W.H. and Reekmans,R.J.  
TITLE Bax-responsive genes for drug target identification in yeast and fungi  
JOURNAL Patent: WO 02064766-A 55 22-AUG-2002;  
JANSSEN PHARMACEUTICA N.V. (BE)  
FEATURES  
source  
1..2027  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4932"  
BASE COUNT 583 a 380 c 427 g 637 t  
ORIGIN  
Alignment Scores: 1.02e-40 Length: 2027  
Pred. No.: 636.00 Matches: 167  
Score: 46.29% Conservative: 70  
Best Local Similarity: 32.62% Mismatches: 197  
Query Match: 22.03% Indels: 78  
Gaps: 14  
US-09-712-338-2\_COPY\_19\_555 (1-537) x AX536454 (1-2027)  
Qy 1 LeuProGlySerThrProAlaSerValGlyArgArgGlnLeu-----ProLys 16  
Db 627 CTTCACAGATACACACAAACCCCTTAATTTGGACCGCTTGAATCAAGATGATCCGCTG 686  
Qy 17 AsnProThrGlyValLysThrLeuThrThrAlaAsnAsnValThrIleArgTyrLysGlu 36  
Db 687 TTACAACTTTTATTTCTCTGTGCACACAGATTACAGTTTGAGACTTAGAACACATAGAT 746  
Qy 37 ProGlyAlaGluGlyValCysLysLysThrThrProGlyValLysSerTyrSerGlyTyrVal 56  
Db 747 CCTTCTAACTAGGAAT-----GACACCGTAAACAAATGTCGGGTACATG 794  
Qy 57 AspThrSerProGluSerHisThrPhePheTrpPheGluAlaArgHisAsnProGlu 76  
Db 795 GACTATAGGATTCACAAACACTTTTACTGCTTTTGAAGTAGGAACGATCCTGCT 854  
Qy 77 ThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeu 96  
Db 855 AACGACCAATATTTCTTTGTTAAATGTTGACCTGGTTCTCTCGTTTACTGGGTG 914  
Qy 97 PheGluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHis 116  
Db 915 CTATTGTAACATAGGCCCTTCATCAATTTGGCGCGGATATGAACCAATCCACAAATCCCTAT 974  
Qy 117 SerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer 136  
Db 975 TCTTGAATTAATACGCTTCATGATCTTCTTAGAACACAGCCACTCGGAGTCGGCTTTCC 1034  
Qy 137 TyrSerAspThrValAspGlySerIleAsnProValThrGlyValGluAsnSerSer 156  
Db 1035 TATGTGAT----- 1043  
Qy 157 PheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrIleIleAspThrThrAsn 176

1044	Db	-----GAAAAAGTCTCCCTCTACAAA	1064
177	QY	LeuAlaAlaGluAlaAlaThrGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeu	196
1065	Db	TTAGCAGCAAGATGCGTACATTTCTCTGGAATGTGTTTGAAGCTTTTCTCATTA	1124
197	QY	AspSerArgValGlnSerLysAspPheSerLeuThrGluSerTyrGlyGlyHisTyr	216
1125	Db	-----CGCTCCACGATTCCACATTCCAGCGCAICTATGAGGAGCAITAT	1172
217	QY	GlyProAlaPhePheAsnHisPheTyrGluGlnAsn---GluArgIleAlaAsnGlySer	235
1173	Db	ATCCCTCAAAFTGCACATGAGATCGTTGTCGAAGAACCTGAAAGA-----	1217
236	QY	ValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGlu	255
1218	Db	-----ACGTTCAAATTAACTTCAGTTATGATTGTTATGATGTTATGCACAGACCT	1265
256	QY	AlaIleGlnAlaProTyrProGluPheAlaVal---AsnAsnThrTyrGlyIleLys	274
1266	Db	TTGATTCAAGCAGATTATTATGACCAATGCGATCGGGAAGGGCGCTATCACCCGT	1325
275	QY	AlaValAsnGluThrValTyrAsnTyrMetIlyspheAlaAsnGlnMetProAsnGlyCys	294
1326	Db	CTCTCATCAGCAAGTAATGTGAATAAGTAAAGTCGACGTCTGCT-----	1370
295	QY	GlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu	314
1371	Db	-----TCTCGTAGCTTGAACAAGTTATGTTATGCTTCTTAATCAAGT	1412
315	QY	CysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrAla	332
1413	Db	TTACATGTCATAGCGGCATGCTTACTGTGACTTGCACITTTGGAAACGTCATTAA	1472
333	QY	PheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro-----	348
1473	Db	---ACAGGACTCAACGTCATGACATAGAGGCGCCTGTGACAGATATAGTACTAGTGT	1529
349	QY	---ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAsp	365
1530	Db	ATCTGTTATACAGCTCTCCGCTATGTCGACAGTATATGAATTTCTGAAGTTCAAGA	1589
366	QY	AlaIleGlyValAsnIle---AsnIlyThrGlnSerAsnAsnAspValTyrTyrAlaPhe	384
1590	Db	ACGCTAGGTCGACGTCGATTAATTCCTGCTGTGATTAACGCTGTTTCAACGGAAT	1649
385	QY	GlnGlnThrGlyAspPheValThrProAsnPheIleGluAspLeuGluIleLeuAla	404
1650	Db	TTGTTTACGGCGCATGAGTAACCA---TTTCAACAATATATGCTGAATTTATAAT	1706
405	QY	LeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGly	424
1707	Db	CACAACATCCGGTATTATATATGCGGGTGAAGGATTATATTTGTAATTTGGCTGGA	1766
425	QY	GlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGly	444
1767	Db	AAACATGCTTGTGCCAAGATGTTGGAATGATCAATAACGTAGGTATCAGAGAAGATG	1826
445	QY	TyrThrPro-----LeuLysValAsnGlyValGluTyrGlyGluThrArgLutyrGly	462
1827	Db	TTAAGACCATGGGTGAGTAAAGAAACACAGGTGAAGAGTTGGACACAGTCAAGACTAT	1886
463	QY	AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle	482
1887	Db	CCTTTCACCTTTTGGAGATATACGATGCGGTCAITATGTCGCCCTATGATCAACCGG	1946
483	QY	AlaSerLeuGlnLeuPheAsnArgThrIlePheGly	494
1947	Db	GCAAGTTTGAATGTCACACAGTTGATTTCCGGT	1982

Search completed: September 17, 2003, 00:48:45

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 16, 2003, 23:02:58 ; Search time 428 Seconds  
(without alignments)  
3086.392 Million cell updates/sec

Title: US-09-712-338-2\_COPY\_19\_555

Perfect score: 2887  
Sequence: 1 LPGSTPASVGRRLPKNPITG.....HTQSSVPLPTATSMSSVGM 537

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1660708 seqs, 1229959015 residues

Total number of hits satisfying chosen parameters: 3321416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-O/Cgn2\_1/USPTO.spool\_p/US09712338/runat\_16092003\_144915\_19194/app\_query.fasta.1.711  
-DB=PublishedApplications\_NA -QRM=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS-human40.cdi -LIST=45 -LOCAL -OUTFMT=ptc -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09712338.cgn.1.1.107 -runat\_16092003\_144915\_19194  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA:

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	588.5	20.4	2218	14	US-10-206-619-1
					Sequence 1, Appli

2	474	16.4	1551	10	US-09-938-842A-683	Sequence 683, App
3	361	12.5	1551	13	US-10-084-018-4	Sequence 4, Appli
4	359	12.4	1697	11	US-09-796-753-39	Sequence 39, Appl
5	359	12.4	2076	10	US-09-909-320-163	Sequence 163, App
6	359	12.4	2076	10	US-09-909-088B-163	Sequence 163, App
7	359	12.4	2076	10	US-09-905-291A-163	Sequence 163, App
8	359	12.4	2076	10	US-09-902-853-163	Sequence 163, App
9	359	12.4	2076	10	US-09-907-841-163	Sequence 163, App
10	359	12.4	2076	10	US-09-907-841-163	Sequence 163, App
11	359	12.4	2076	11	US-09-904-011-163	Sequence 163, App
12	359	12.4	2076	11	US-09-906-742-163	Sequence 163, App
13	359	12.4	2076	11	US-09-906-838-163	Sequence 163, App
14	359	12.4	2076	11	US-09-907-613-163	Sequence 163, App
15	359	12.4	2076	11	US-09-907-942-163	Sequence 163, App
16	359	12.4	2076	11	US-09-904-859-163	Sequence 163, App
17	359	12.4	2076	11	US-09-909-204-163	Sequence 163, App
18	359	12.4	2076	11	US-09-904-820-163	Sequence 163, App
19	359	12.4	2076	11	US-09-904-786-163	Sequence 163, App
20	359	12.4	2076	11	US-09-906-646-163	Sequence 163, App
21	359	12.4	2076	11	US-09-906-700-163	Sequence 163, App
22	359	12.4	2076	11	US-09-903-786-163	Sequence 163, App
23	359	12.4	2076	11	US-09-902-903-163	Sequence 163, App
24	359	12.4	2076	11	US-09-903-749A-163	Sequence 163, App
25	359	12.4	2076	11	US-09-904-119-163	Sequence 163, App
26	359	12.4	2076	11	US-09-904-956-163	Sequence 163, App
27	359	12.4	2076	11	US-09-902-736-163	Sequence 163, App
28	359	12.4	2076	11	US-09-907-794-163	Sequence 163, App
29	359	12.4	2076	11	US-09-903-943-163	Sequence 163, App
30	359	12.4	2076	11	US-09-904-462-163	Sequence 163, App
31	359	12.4	2076	11	US-09-907-925-163	Sequence 163, App
32	359	12.4	2076	11	US-09-902-692-163	Sequence 163, App
33	359	12.4	2076	11	US-09-903-520-163	Sequence 163, App
34	359	12.4	2076	11	US-09-905-056-163	Sequence 163, App
35	359	12.4	2076	11	US-09-909-064-163	Sequence 163, App
36	359	12.4	2076	11	US-09-904-553-163	Sequence 163, App
37	359	12.4	2076	11	US-09-905-381-163	Sequence 163, App
38	359	12.4	2076	11	US-09-905-088-163	Sequence 163, App
39	359	12.4	2076	11	US-09-907-575-163	Sequence 163, App
40	359	12.4	2076	11	US-09-905-075-163	Sequence 163, App
41	359	12.4	2076	11	US-09-902-759-163	Sequence 163, App
42	359	12.4	2076	11	US-09-902-634-163	Sequence 163, App
43	359	12.4	2076	11	US-09-902-713-163	Sequence 163, App
44	359	12.4	2076	11	US-09-907-979-163	Sequence 163, App
45	359	12.4	2076	11	US-09-902-615-163	Sequence 163, App

ALIGNMENTS

RESULT 1  
US-10-206-619-1  
: Sequence 1, Application US/10206619  
: Publication No. US20030108908A1  
: GENERAL INFORMATION:  
: APPLICANT: Korea Research Institute of Bioscience and Biotechnology  
: APPLICANT: Dong Kook Pharmaceutical Co.  
: APPLICANT: Rhee, Sangki  
: APPLICANT: Choi, Eulsung  
: APPLICANT: Kang, Hyunah  
: APPLICANT: Sohn, Junghoon  
: APPLICANT: Bae, Junghoon  
: APPLICANT: Kim, Moowoon  
: APPLICANT: Agaphonov, Michael  
: TITLE OF INVENTION: Hansenula polymorpha mutants and process for the preparation o  
: TITLE OF INVENTION: recombinant proteins using the same  
: FILE REFERENCE: 4220-116 US  
: CURRENT APPLICATION NUMBER: US/10/206,619  
: CURRENT FILING DATE: 2002-07-26  
: PRIOR APPLICATION NUMBER: US/09/674,617  
: PRIOR FILING DATE: 2001-01-03  
: PRIOR APPLICATION NUMBER: PCT/KR00/00173  
: PRIOR FILING DATE: 2000-03-04  
: NUMBER OF SEQ ID NOS: 9  
: SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 2218

; TYPE: DNA

; ORGANISM: Hansenula polymorpha DL1

; PUBLICATION INFORMATION:

; DATABASE ACCESSION NUMBER: U67174

; DATABASE ENTRY DATE: 1996-08-17

; RELEVANT RESIDUES: (1)..(2218)

US-10-206-619-1

Alignment Scores:

Pred. No.: 1,17e-59 Length: 2218

Score: 588.50 Matches: 169

Percent Similarity: 45.61% Conservative: 70

Best Local Similarity: 32.25% Mismatches: 200

Query Match: 20.38% Indels: 86

DB: 14 Gaps: 20

US-09-712-338-2\_COPY\_19\_555 (1-537) x US-10-206-619-1 (1-2218)

QY 1 LeuProGlySerThrProAlaSerValGlyArgGlu-----Leu 14  
 DB 640 ATACCTCAGAACATGAGCTCAGTCAGCAGCGGCGAGTCAAAATTACAAAGAACCTG 699  
 QY 15 ProLysAsnProThrGlyValLysThrLeuThrThrAlaAsnAsnValThrIleArgTyr 34  
 DB 700 CCGATTTTGGGACTTCAATGTTCTCAATGAGAAGTTC-TCCAACTACAAGCTGAGGGTT 758  
 QY 35 Lys-----GluProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyr 52  
 DB 759 AAGAAGACCGGAGCGGAGCATTTGGGACTGGAGCCACACA-----AGACAGTAC 806  
 QY 53 SerGlyTyrValAspThrSerProGlu---SerHisThrPhePheThrPheGluAla 71  
 DB 807 TCGGATACTGTGATGGAGGAGGAGACAGACATTTCTTATGGATGTTGAGTCC 866  
 QY 72 ArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyProGlySerAsp 91  
 DB 867 AGAAATGACCCGGTCAACGACCCCTGTGATTCGTGGCTCAACGGTGTCCAGGATGCTCT 926  
 QY 92 SerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsnSerThrPheAspAsp 111  
 DB 927 TCCTTGACTGGAATGCTTTTGGAGCTCGGCTCTGCTTATCGGTCCAGATCTCAAGCCA 986  
 QY 112 TyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeu 131  
 DB 987 ATCAACACCCATATTCGTGAATTCCAATGCCACCTGTGATTTCCCTTGACCGCCGTG 1046  
 QY 132 GlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnProValThrGlyVal 151  
 DB 1047 AATGTTGGATTCTCGTACTCT----- 1067  
 QY 152 ValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeu 171  
 DB 1068 -----TCCAGTCT 1076  
 QY 172 IleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyLeuSer 191  
 DB 1077 GTTTCCTAACAGGTCGCGAGCTGGTAAGAGCTCTATGCTTCTTGGAGTTGTTCTTACCAG 1136  
 QY 192 GlyLeuProSerLeuAspSerArgValGlnSerLysAspPheSerLeuThrThrGluSer 211  
 DB 1137 CAATTCACCACTTG-----CTGAAGAAGACACTTCCACATCGCGCGGAGTCG 1184  
 QY 212 TyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGluGlnAsnGluArgIle 231  
 DB 1185 TACGGTGTGCTATACATCCCACTGTTGCCCTCCAGATTCACCCATGCTGACAGA--- 1241  
 QY 232 AlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGly 251  
 DB 1242 -----TCTTTCACCCCTCACTTCGGTGTGTTGATGTGAACGCT 1277  
 QY 252 IleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAlaValAsnThrTyr 271

DB 1278 TTGACCGACCCACTTAACCAAGTACCCTTCTACGAGAGAAATGCATGCTCTACTGATGGT 1337  
 QY 272 GlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetPro 291  
 DB 1338 GGCTATGACGAA---CCCTGGCAGCTCTGAGTCGGAAGAAATGTTGGAGACCTTGCTT 1394  
 QY 292 AsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAsp 311  
 DB 1395 AGA---TGTTGTCAATTGATTGAATCATGCTACAGCTCGCAG-----TCT 1436  
 QY 312 TyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyr 331  
 DB 1437 GTGTTCTCATGTCTCCCGGCTCCATCTACTGCAACACGACCAACTTGACCATTCCAA 1496  
 QY 332 AlaPheAlaGlyArgGlyValTyrAspIleArgHisPro-----Tyr 345  
 DB 1497 AAG---ACCGGCAGAAACGCTCTACGAGCTTAGAAGATGTCGAGGGAACCTCTGTGCTAC 1553  
 QY 346 AspAspProThrProSerTyrProSerTyrAsnLysPheLeuAlaLysAspSerValMetAsp 365  
 DB 1554 AAAGAC-----ATGGAATACATTGACCAATATTGACCAAGGACTTGTCAAGGAA 1604  
 QY 366 AlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAspValTyrTyrAlaPhe 384  
 DB 1605 AAGGTTGGCGCTCAGGTTGACACTTACGAGTCGTGTATTTCCGACGTGAACAGAACTTC 1664  
 QY 385 GlnGlnThrGlyAspPheValTrpPro-----AsnPheIleGluAspLeuGluGlu 401  
 DB 1665 CTGTTTCTGCTGATGGATGAAACCTTACCACAGAACGTTATCAATCTCTGCGAGCAA 1724  
 QY 402 IleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAlaAspTyrIleCys 420  
 DB 1725 -----GGTCTCTCTGTC-----CTGATTTACGCGAGGAGCAAGGATTTTCATCTGC 1769  
 QY 421 AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGlnPhe 440  
 DB 1770 AATGGCTCGGAACCAAGCTGCTCCATGAGCTCCCTGGTCTGGACACGATGAATTC 1829  
 QY 441 ArgSerAlaGly---TyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArg 459  
 DB 1830 GAGTCCCCCGAGCTGTACAACTCCCTGTAAGGATGGCCTAAGTGGCGAGGTCAAG 1889  
 QY 460 GluTyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGlnValProTyrTyr 479  
 DB 1890 AATGCTGGCAAGTTCACTTTCGTAAGATGTTTGTGAGGAGACACATGGTTCCATACGAC 1949  
 QY 480 GlnProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTyrAspIleAlaGlu 499  
 DB 1950 CAGCTGAGAGCTCTTTGGCTATGTCATATGATGATAGCTGGT---GACTACTCTCTG 2006  
 QY 500 GlyGlnLysLys 503  
 DB 2007 GGAACCAAGAAA 2018

## RESULT 2

US-09-938-842A-683

; Sequence 683, Application US/09938842A

; Patent No. US20020160378A1

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff

; APPLICANT: Kreps, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; FILE REFERENCE: SCRIPL300-3

; CURRENT APPLICATION NUMBER: US/09/938,842A

; CURRENT FILING DATE: 2001-08-24

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/227,866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264,647

; PRIOR FILING DATE: 2001-01-16



; PRIOR APPLICATION NUMBER: US 60/300,111  
 ; PRIOR FILING DATE: 2001-06-22  
 ; NUMBER OF SEQ ID NOS: 5379  
 ; SEQ ID NO 683  
 ; LENGTH: 1551  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 ; US-09-712-338-2\_COPY\_19\_555 (1-537) x US-09-712-338-2\_COPY\_19\_555 (1-1551)

## Alignment Scores:

Pred. No.: 3,31e-46 Length: 1551  
 Score: 474.00 Matches: 141  
 Percent Similarity: 45.02% Conservative: 67  
 Best Local Similarity: 30.52% Mismatches: 174  
 Query Match: 16.42% Indels: 80  
 DB: 10 Gaps: 18

US-09-712-338-2\_COPY\_19\_555 (1-537) x US-09-712-338-2\_COPY\_19\_555 (1-1551)

```

Qy 45 ThrThrProGlyValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
Db 277 ACTGTCGACGATTTAGTCATCATCGCTGCTTATACAGCTCCCGAATCTCGTGGCGCA 336
Qy 63 HisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
Db 337 AGCATGTTCTACTTCTTCTCGATCACGG---AACAAAGAGGAGTCTCGTGTGTGATT 393
Qy 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 102
Db 394 TGGITGCGGAGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 453
Qy 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 122
Db 454 TTCAAGATCACTAGTACATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 513
Qy 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
Db 514 AATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 573
Qy 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
Db 574 GATATCGGTCATGACGAACTGGAGIT-----600
Qy 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAla 182
Db 601 -----AGCATGATCTT 612
Qy 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
Db 613 TATGATTTTCTGACGCTTCTTTGCGGAGCACCCTAAGTTG-----GCAGAAA 660
Qy 203 LysAspPheSerLeuThrPheGluSerTyrGlyHisTyrGlyProAlaPhePheAsn 222
Db 661 AACGCTTTTATATTACTGGAGTATCATGCTGGGCACATATATCCAGAGCTTTTGCTTC 720
Qy 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
Db 721 CGAGTCCATAAAGGAAACAG-----GCTAAT-----CAGGAGTTCATATAAC 765
Qy 243 PheAsnSerLeuGlyIleLeuAsnGlyIleAspGluAlaIleGlnAlaProTyrTyr 262
Db 766 CTAAGGGATTCGCTATATGGAATGGCTTACAGATCTCGACATCCCAATACCCAGCCTAT 825
Qy 263 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 282
Db 826 CTGACTACGCTTTGGAA-----ATGGGTCTAATCACGCAAAAGAG-----867
Qy 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
Db 868 CATGATCGCTTAGAAAGATTGTCACATG---TGCAGACTATCAATCAAGCTTTGTGGA 924
Qy 303 GlnThrAsnArgThrAla---LeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMet 321
Db 1551
  
```

```

Db 925 ACTGATGCGACAACTTCTTGTTGCAATCATATCTTGCTGCAACTCTTGTTCAGTGGT 984
Qy 322 CysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIle 341
Db 985 GTAATGAGTCACTGCTGGTGGATTAACATAT-----TACGACATC 1023
Qy 342 ArgHisPro-----TyrAspAspProThrProProSerTyrTyrAsn 355
Db 1024 AGGAAGAAGTCGTTGGGAAGTTGTGCTACGATTTC-----TCAACATGGAG 1071
Qy 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyVal---AsnIleAsnTyrThr 374
Db 1072 AAATTTTGAATCTCCATCTGTCGAGGAGTCACITGCTGTTGGGACATAGACTTTGTC 1131
Qy 375 GlnSerAsnAspValTyrTyrAlaPheGlnGlnThrGlyAspPheValTyrProAsn 394
Db 1132 TCCTGCAGTACTTCACTATCAGTCTATCAGGCAATGCTTGTAGAT-----IGGATGAGG 1179
Qy 395 PheIleGluAsp-----LeuGluGluIleLeuAlaLeuProValArgVal 409
Db 1180 AATCTGAGGTTGGAATTCACGCTCTTGGAAATGAATCAGCCTC-----1227
Qy 410 SerLeuIleTyr---GlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyAlaVal 428
Db 1228 ---CTTGTGTAATGTCGAGATATGATCTTATCTGCAACTGCTGCTGTAACTCGAGTGG 1284
Qy 429 SerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeu 448
Db 1285 GTGAATGCAATGAGTGTGTCAGGGAACAACTTTGGGGCGCTAAAGAAAGTTCCGTTT 1344
Qy 449 LysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArg 468
Db 1345 ATTGTGATGCAAGAAAGCAGGCTGTTAAAGACTTATGAACAACATCAGTTTCTCTCAAG 1404
Qy 469 ValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnPhe 488
Db 1405 GIGAGAGATGACGACACATCTGTTCCATGACGACGCTAAAGTGCATTTGAAATGCG 1464
Qy 489 AsnArg 490
Db 1465 AAGCGA 1470
  
```

## RESULT 3

US-10-084-018-4  
 ; Sequence 4, Application US/10084018  
 ; Publication No. US20020160499A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bandman, Olga  
 ; Hawkins, Phillip R.  
 ; Hillman, Jennifer L.  
 ; Lal, Preeti  
 ; Goli, Surya K.  
 ; TITLE OF INVENTION: NOVEL HUMAN SERINE  
 ; CARBOXYPEPTIDASE  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/084,018  
 ; FILING DATE: 25-Feb-2002  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/299,689A  
 ; FILING DATE: <Unknown>

```

; APPLICATION NUMBER: 08/828,488
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0241 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1551 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: MMLR3DT01
; CLONE: 566993
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-084-018-4

Alignment Scores:
Pred. No.: 1.09e-32 Length: 1551
Score: 361.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.50% Indels: 116
DB: 13 Gaps: 21

US-09-712-338-2_COPY_19_555 (1-537) x US-10-084-018-4 (1-1551)
QY 47 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 281 CCAGGACTGAACATGAAGAGTATGTCGGCTTCCTCACCGTGAATAGACATACACAGC 340
QY 63 HisThrPhePheTrpPhePheGluAlaAArgHisAsnProGluThrAlaProIleThrLeu 82
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 341 AACCTCTTCCTCTGGTCTTCCCGACTCAGATACAGCCAGAGATGCCCCAGTACTGTTCTC 400
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 102
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 401 TGGCTACAGGGTGGCGGGAGGTTCATCCATGTTWGCACCTTTGTGGAACATGGGCT 460
QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 461 TATGTTGTCAACAGTAACATGACCTTGGGTGCACAGAGACTTCCCTCGGACCACACAGSTC 520
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 521 TCCATGCTTTACATTTGACATCCAGTGGGCACAGCTTCAGTTTACTGATGATACCCAC 580
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 581 GGA----- 583
QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAla 182
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 584 -----TATGCAGTCAATGAGCAGCATGTATGACCGGATTTATAC 622
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 623 AGTCACCTAATTCAGTTTTC-----CAGATATTTCCCTCAATATAAAT 667
QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsn 222
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 668 AATGACCTTTTATGTCACATGGGAGCTTTATGCAGGAAATAATGTGCCACCATTCACAC 727
QY 223 HisPheTyrGluGlnAsnGluAArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 728 CTCATCCATCCCTCAAC-----CCTGTGAGAGAGGTGAGATCAAC 769
QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyr 262
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 39
; LENGTH: 1697
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)...(1528)
US-09-796-753-39

```

## Alignment Scores:

```

Pred. No.: 2,18e-32 Length: 1697
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.44% Indels: 116
DB: 11 Gaps: 21

```

US-09-712-338-2\_COPY\_19\_555 (1-537) x US-09-796-753-39 (1-1697)

```

QY 47 ProGly-----ValIysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 299 CCAGGACTGAACGTGAAGAGTATGCGCGCTTCCCTCACCGTGAATAGACTTACACAC 358
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 63 HisThrPhePhePhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 359 AACCTCTTCTTCTGTTCTTCCACGCTCAGATACAGCCAGAGAGTGCCTTCTC 418
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGlyPro 102
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 419 TGGCTACAGGGTGGCGGAGGTTCATCCATGTTGTCGACTCTTTGTGAACATGGCCT 478
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 122
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 479 TATGTTGTCACAAAGTAACTGACCTGGGTGACAGAGACTTCCCTGACCAACACGCTC 538
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 ICCATGCTTTACATTCACATCCAGTGGCAGAGCTTCAGTTTCTAGTATGATACCAC 598
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 599 GGA----- 601
QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla 182
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 602 -----TATGCAGTCAATGAGGACGATGATAGCAGCGGATTTATAC 640
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 641 AGTCGACTAATTCAGTTTTC-----CAGATATTCCTGAAATATAAAAT 685
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsn 222
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 686 AATGACTTTTATGTCAGTGGGAGCTTATGTCAGGAAATATGTCAGCCATTCACAC 745
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 746 CTCATCCATTCCTCCCAAC-----CCTGTGAGAGAGGTCAGATCAAC 787
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleAspGluAlaIleGlnAlaProTyrTyr 262
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 788 CTGACGGAATTCCTATGAGAGTGGATATTCGATCCGAAATCAATATAGGGGGCTAT 847
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleIleValAsnGlnGluThrValTyrAsn 282
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 848 GCAGAAATC-----CTGTACCAAAATGGCTTGGATGAGAGCAAAAG 895
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 896 TAC-----TTCAGAGACAG-----TGCATGAATGCATAGAACATCAGG 937
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 938 AAGCAGAACTGGTTGAGGCTTTGAATACTGGATAAACTACAGATGCGGACTTAACA 997
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 998 AGTGATCTCTTCTTACTTCCAGAAATGTACAGGATAGTAAATTAATACTTATG--- 1054
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 355
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1055 -----CGGTGCACGGAACCTGAGGATCAGCATCAGCTTTACTATGTG 1090
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1091 AAATTTTGTCACTCCAGAGGTGAGACAGCCATCCACGTGGGG---AATCAGACTTTT 1147
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 376 SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTyrPro 393
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1148 AATGATGGAACATAGTGTGAAAGAGTACTTGGAGAGAGATACAGTACAGTCAGTTAAGCCA 1207
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 394 AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1208 TGGTTAACTGAATCAATGAATAAT-----TATAAGGTCTCTGATACAAT 1252
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 414 GlyAspAlaAspTyrIle----- 421
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1253 GGCACCACTGACATCATCTGCGAGCTGCCCTGCAGAGCGCTCCTGTGATGGCATGGAC 1312
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

QY 422 trpPheGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 439
Db 1313 TGAAGAAGTCCAG-----GAAACAAAGGACGCAAAAAAGTT 1354
QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452
Db 1355 TGAAGATCTTAAATCTGACAGTGAAGTGGCTGGTTAC----- 1393
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
Db 1394 -----ATCGGCAAGCGGTGACTTCCATCAGCGTAATATTCGAGGTGGA 1438
QY 473 GlyHisGluValProTyrGlnProIleAlaSerLeuGlnPheAsnArgThrIle 492
Db 1439 GGACATATTTACCTATGACACAGCCTCGACAGCTTTGACATGATTATCGATTCAAT 1498
QY 493 Phe-----GlyTrpAsp 496
Db 1499 TATGGAAGGATGGAT 1516

RESULT 5
US-09-909-320-163
; Sequence 163, Application US/09090320
; Patent No. US20020132240A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,320
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089

```

```

; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-909-320-163

Alignment Scores:
Pred. No.: 2,99e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.44% Indels: 116
DB: 10 Gaps: 21

US-09-712-338-2_COPY_19_555 (1-537) x US-09-909-320-163 (1-2076)
QY 47 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
Db 295 CCAGGACTGACATGAGAGTTATCGCGCTTCCTCACCCTGATATACACACAGC 354
QY 63 HisThrPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
Db 355 AACCTCTCTCTGTTCTTCCAGCTCAGATACAGCAGAGAGATGCCACAGTATGTC 414
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuLeuGlyLeuGluLeuGlyPro 102
Db 415 TGGCTACAGGTGGCGCGGAGGTTCATCCATGTTTGGACTCTTTTGGAAACATGGGCT 474
QY 103 CysHisValAsnSerThrPheAspTyrIleAsnProHisSerTyrAsnGluValSer 122
Db 475 TATGTTGTCAAGTACATGACCTTGGTGACAGACATCCCTGGACCACACGCTC 534
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
Db 535 TCCATGCTTTACATTCACATCCAGTGGGACAGGCTTCAGTTTACTGATGATACCCAC 594
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
Db 595 GGA-----TATCCAGTCAATGAGGACGATGTAGCAGCGGATTTATAC 636
QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla 182
Db 598 -----TATCCAGTCAATGAGGACGATGTAGCAGCGGATTTATAC 636
QY 183 TrpGluLeuLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
Db 637 AGTGCACATAATTCAGTTTTC-----CAGATATATTCTCTGGAATATAAAAA 681
QY 203 LysAspPheSerLeuThrThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsn 222
Db 682 AATGACTTTTATGACTGGGAGTCTTATCCAGGGAATATGTCACAGCCATTCACAC 741
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
Db 742 CTCATCCATTCCCTCAAC-----CCTGTGAGAGAGTGAAGATCAAC 783

```

```

QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 262
Db   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
784 CTGAACGGAATGCTATTGGAGATGGATATTCGATCCCGAATCAATTAAGGGGGCTAT 843
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleIleValAsnGluThrValTyrAsn 282
Db   ||||||  |||   |||   |||   |||   |||   |||   |||   |||   |||
844 GCAGAAATTC-----CTGTACCAAAATGGCTGTGGATGAGAACCAAAAAAG 891
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
Db   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
892 TAC-----TTCAGAACGAG-----TGCATGAATCATAGAACACATCAGG 933
QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
Db   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
934 AAGCAGAACTGGTTGGAGCCTTTGAAIACTGATAAAGTACTAGATGGCGACTTACCA 993
QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335
Db   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
994 AGTATCCTCTTACTTCCAGATGTTACAGGATGTAGTAATTAATACTTTTGG--- 1050
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 355
Db   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
1051 -----CGGTGCAGGAACCTGAGGATCAGCTTACTATGTG 1086
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
Db   |||||||  |||   |||   |||   |||   |||   |||   |||   |||   |||
1087 AAATTTTGTCACTCCAGAGGTGAGACAAGCCATCCACGTGGGG---AATCAGACTTT 1143
QY 376 SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValIrrPro 393
Db   |||||||  |||   |||   |||   |||   |||   |||   |||   |||   |||
1144 AATGATGGAAGTACTGTGTAAGTAAATCTTCGAGAGATACATACAGTACAGTAAAGCCA 1203
QY 394 AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413
Db   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
1204 TGGTTRACTGAATCATGAATAAT-----TATAAGGTTCGATCTACAAAT 1248
QY 414 GlyAspAlaAspTyrIle-----CysAsn 421
Db   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
1249 GGCACACTGGACATCATCGTGGCAGCTGCCCTGCACAGAGCGCTCTGTATGGCGATGGAC 1308
QY 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 439
Db   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
1309 TGGAAAGGATGCCAG-----GAATACAAAGAGGAGCAAAAAAAGTT 1350
QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452
Db   |||||||  |||   |||   |||   |||   |||   |||   |||   |||   |||
1351 TGAAGATCTTTAAATCTGCACATGAGTGGCTGGTTAC----- 1389
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValIlyrGluAla 472
Db   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
1390 -----ATCCGGCAAGCGGTGACTTCCATCAGTAATTAATTCAGGTGGA 1434
QY 473 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492
Db   |||||||  |||   |||   |||   |||   |||   |||   |||   |||   |||
1435 GCACATATTTTCCCTATGACAGCGCTCTGAGAGCTTTTGACATGATTAATCGATTCATT 1494
QY 493 Phe-----GlyTrpAsp 496
Db   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
1495 TATGGAAGGATGGGAT 1512

```

## RESULT 6

```

US-09-909-088B-163
; Sequence 163, Application US/09909088B
; Patent No. US20020146709A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen

```

```

; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,088B
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-909-088B-163

```

## Alignment Scores:

```

Pred. No.: 2,99e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.44% Indels: 116
DB: 10 Gaps: 21

```

US-09-712-338-2\_COPY\_19\_555 (1-537) x US-09-909-088B-163 (1-2076)



US-09-712-338-2\_COPY\_19\_555 (1-537) x US-09-905-291A-163 (1-2076)

US-09-902-853-163  
; Sequence 163, Application US/09902853  
; Publication No. US20020192659A1  
; GENERAL INFORMATION:

APPLICANT: Genentech, Inc.  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gieritsen, Mary E.  
APPLICANT: Goddard, A.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kijavini, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: 10466-14  
CURRENT APPLICATION NUMBER: US/09/902,853  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: US/09/665,350  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: PCT/US99/28214  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30  
PRIOR APPLICATION NUMBER: PCT/US99/28564  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/28565  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30095  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: PCT/US99/30911  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US99/30999  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US00/00219  
PRIOR FILING DATE: 2000-01-05  
NUMBER OF SEQ ID NOS: 423  
SEQ ID NO 163  
LENGTH: 2076  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-09-902-853-163

Alignment Scores: 2.99e-32 Length: 2076  
Pred. No.:

Score: 359.00 Matches: 130  
Percent Similarity: 39.51% Conservative: 62  
Best Local Similarity: 26.75% Mismatch: 178  
Query Match: 12.44% Indels: 116  
DB: 10 Gaps: 21  
US-09-712-338-2\_COPY\_19\_555 (1-537) x US-09-902-853-163 (1-2076)  
QY 47 ProGly-----VallySerTyrSerGlyTyr-----ValAspThrSerProGluSer 62  
DB 295 CCAGGACTGAACATGAAGATTATCGCGGCTTCTCAGCGTGAATAAGACTTACAACAGC 354  
QY 63 HisThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82  
DB 355 AACCTCTCTCTCGTTCTTCCAGCTCAGATACAGCAGACAGATGCCAGTAGTATCTC 414  
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuPro 102  
DB 415 TGGCTACAGGTTGGCGGGAGGTTTCATCCATGTTGGACTCTTTGGACATGGGCTT 474  
QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 122  
DB 475 TATGTTGTCAACAAGTAACATGACCTTGCSTGCAGAGACTTCCCTGGACACACAGCTC 534  
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValASP 142  
DB 535 TCCATGCTTTACATTCACATCCAGTGGCAGCGCTTCAGTTTTCATGATGATACCCAC 594  
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162  
DB 595 GGA----- 597  
QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAla 182  
DB 598 -----TATGCAGTCAATGAGCAGCATGTAGCAGCGGATTTATAC 636  
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202  
DB 637 AGTGCACATAATTCAGTTTTC-----CAGATATTTCTCGAATATAAANAAT 681  
QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsn 222  
DB 682 AATGACITTTATGTCACCTGGGACTCTTATCCAGGGAATATGTGCCAGCATATGCACAC 741  
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242  
DB 742 CTCATCCATTCCTCAAC-----CCTGTGAGAGAGTGAGATCAAC 783  
QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 262  
DB 784 CTGAACGGAATGCTATTGGAGATGGATATCTCTCCGAATCAATATAGGGGGGTAT 843  
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 282  
DB 844 GCAGAAATC-----CTGACCAATTCGTTGTGGATGAGAGCAAAAAAG 891  
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302  
DB 892 TAC-----TCCAGAAAGCAG-----TGCATGAATCATAGACACATCAGG 933  
QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319  
DB 934 AAGCAGAACTGGTTTGGAGCGCTTCAAAATCTGATAAATCTAGATAGCGGCTTACA 993  
QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335  
DB 994 AGTGATCCTTCTTACTTCCAGAGATTACAGAGATGTAGTAATTAATACTTTTGTG--- 1050  
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 355  
DB 1051 -----CGGTGACGAGAACCTGAGGTACAGCTTACTATGTG 1086  
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375



```

Db 1087 AAATTTTGTCACTCCAGAGGTGACAGCAAGCCATCCAGTGGGG---AATCAGACTTTT 1143
QY 376 SerAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTrpPro 393
Db 1144 AATGATGAAGTATAGTTGAAAGTACTTCGGAGAAGATACAGTACAGTACAGTTAAGCCA 1203
QY 394 AsnPheLeuGluAspLeuGluGluLeuAlaLeuProValArgValSerLeuIleTyr 413
Db 1204 TGGTTAACTGAATCAAGTAATANI-----TATAAGGTTCTGATCTACAAT 1248
QY 414 GlyAspAlaAspTyrIle-----CysAsn 421
Db 1249 GGCAACTGACATCAATCTGGCGAGCTGCCCTGACAGAGCGCTCCITGATGGGCGATGGAC 1308
QY 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 439
Db 1309 TGGNAAGATCCAG-----GAATACAGAAGGCGAGAAAAAAGTTT 1350
QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452
Db 1351 TGGGAAGATCTTAAATCTGACAGTGAAGTGGCTGGTTAC----- 1389
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
Db 1390 -----ATCCGCGCAAGCGGTGACTTCCATCAGGTAATATTTCGAGGTGGA 1434
QY 473 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492
Db 1435 GGACATATTTCACCTATGACCCCTCTGAGAGCTTTTGACATGATTAATCGAATCATT 1494
QY 493 Phe-----GlyTrpAsp 496
Db 1495 TATGGAAAAGGATGGGAT 1512

```

## RESULT 9

```

US-09-907-824-163
; Sequence 163, Application US/09907824
; Publication No. US20020197671A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,824
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414

```

```

; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-907-824-163

Alignment Scores:
Pred. No.: 2,99e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.44% Indels: 116
DB: 10 Gaps: 21

US-09-712-338-2_COPY_19_555 (1-537) x US-09-907-824-163 (1-2076)
QY 47 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
Db 295 CCAGGACTGAACATGAAGAGTATTCGCGGCTCTCCACCGTGAATAAGACTTACAACAGC 354
QY 63 HisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
Db 355 AACCTCTCTCTCTGGTCTTCCCGAGCICAGATACAGCCAGAGATGCCAGATGATGTC 414
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGlyPro 102
Db 415 TGGCTACAGGTGGCGGGAGGTTCATCCATGTTGGACTCTTTGGGAACATGGCCCT 474
QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 122
Db 475 TATGTTGTCAACAAGTACATGACCTTGGGTGACAGAGACTTCCCTGGACCAACAGC 534
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
Db 535 TCCATGCTTATACATGACATCCAGTGGGACAGGCTTCAGTTTACITGATGATACCCAC 594
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
Db 595 GGA----- 597

```

```
QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAla 182
Db 598 -----TATGCAGTCAATGAGGACGATGATGACGGGATTTATAC 636
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
Db 637 AGTGCACATTAATCAGTTTTC -----CAGATAATTCCTGAATATAAAAT 681
QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsn 222
Db 682 AATGACTTTTATGCTACTGGGAGTCTTATGACGGGAAATATGTCGCCCAATTCACAC 741
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
Db 742 CTCATCCATCCCTCAAC -----CCTGTGAGAGAGGAGAGATCAAC 793
QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 262
Db 784 CTGACGGAATTCCTATGAGATGGATATCTGATCCCAATCAATATATAGGGGGCTAT 843
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleIleValAsnGluThrValTyrAsn 282
Db 844 GCAGAAATC -----CTGTACCAAAATGGCTTGTGGATGAGAGCAAAAAAG 891
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
Db 892 TAC -----TTCAGAGACAG -----TGCATCAATCATAGACACATCAGG 933
QY 303 GlnThrAsn -----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
Db 934 AAGCAGAACTGGTTGAGGCTTTGAAATCTGGTAACTAGTAACTAGTGGCGCTTAACA 993
QY 320 Asn -----MetCysArgAspAsnValGluGly -----ProTyrTyrAlaPheAlaGly 335
Db 994 AGTATGCTTCTTACTTCCAGATGTTACAGGATGTAGTAACTACTATACTTTTG --- 1050
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 355
Db 1051 -----CGGTGCAGCAAGCTCAGGATCAGCTTACTATGTG 1086
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
Db 1087 AAATTTTGTCTCCAGAGGTGAGACAGCCATCCAGTGGGG ---AATCAGACTTTT 1143
QY 376 SerAsnAsnAspValIyr -----TyrAlaPheGlnGlnThrGlyAspPheValTrpPro 393
Db 1144 AATGATGGAATVAGTTGAAAGTACTTCCGAGAGATACAGTACAGTACAGTAAAGCCA 1203
QY 394 AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413
Db 1204 TGGTTAACTGAATCAATCAATAAT -----TATAAGGTTCTGATCTACAAT 1248
QY 414 GlyAspAlaAspIyrIle -----CysAsn 421
Db 1249 GGCAACTGGACATCATCTCGTGGCAGCTGCCCTGACAGAGCGCTCCTTGATGGGATGGAC 1308
QY 422 TrpPheGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGln ----- 439
Db 1309 TGGAAAGATCCCAAG -----CAATACAAGAGGAGCAAAAAAGTT 1350
QY 440 -----PheArgSer -----AlaGlyTyrThrProLeuLysValAsnGly 452
Db 1351 TGGAGATCTTTAAATCTGACAGTGAAGTGGCTGGTTAC ----- 1389
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
Db 1390 -----ATCCGCAAGCGGGTACTTCCATCAGTAAATATTTCGAGGTGGA 1434
QY 473 GlyHisGluValProTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492
Db 1435 GGACATATTTTACCTATGACCAAGCCCTGAGAGCTTTTGACATGATTAATCGATTCATT 1494
```

```
QY 493 Phe-----GlyTrpAsp 496
Db 1495 TATGGAAAGGATGGGAT 1512
```

## RESULT 10

```
US-09-907-841-163
; Sequence 163, Application US/09907841
; Publication No. US20020198366A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin B.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,841
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-907-841-163
```

```
Alignment Scores: 2.99e-32 Length: 2076
Pred. No.: 359.00 Matches: 130
Score: 39.51% Conservative: 62
Percent Similarity: 26.75% Mismatches: 178
Best Local Similarity: 12.44% Indels: 116
Query Match:
```

```

DB:          10          Gaps:          21
US-09-712-338-2_COPY_19_555 (1-537) x US-09-907-841-163 (1-2076)
QY 47 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
DB 295 CCAGGACTGCACATGAAGAGTATGCCGGCTTCCTCCACCGTGAATAAGACTTACACAGC 354
QY 63 HisThrPhePhePhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
DB 355 AACCTCTCTCTCTGCTTCCAGCTCAGATACAGCCAGAGATGCCCCAGATGCTTC 414
QY 83 TrpLeuAsnGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 102
DB 415 TGGCTACAGGTGGCGGGAGGTTCATCCATGTTGGACTCTTTGTGACATCGGGCT 474
QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 122
DB 475 TATGTTGTCAACAAGTACATGACCTTGGTGACAGAGACTTCCCTCGACCAACAAGCTC 534
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerIleSerAspThrValAsp 142
DB 535 TCCATGCTTTACATGACATCCAGTGGCAGAGCTTCAGTTTACTGATGATACCCAC 594
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
DB 595 GGA----- 597
QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAla 182
DB 598 -----TATGCAGTCAATGAGGACGATGAGCAGCGCATTTATAC 636
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGluSer 202
DB 637 AGTCCACTAATTCAGTTTC-----CAGATATTCCTGAATATAAAAT 681
QY 203 LysAspPheSerLeuThrTrpGluSerTyrGlyGlyHisTyrGlyProAlaPheAsn 222
DB 682 AATGACTTTTATGTCATCGGGAGTCTTATGCAGGAAATATGTGCCAGCCATTCACAC 741
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
DB 742 CTCATCCATCCCPAAC-----CCTGTGAGAGAGGTGAAGATCAAC 783
QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 262
DB 784 CTGAACGAATGCTATTCGAGATGATATTCATCCCAATCAATATAGGGGGTAT 843
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 282
DB 844 GCAGAAATC-----CTGTACCAAAATGGCTTGTGGATGAGAGCAAAAG 891
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspIleIleSerThrCysLys 302
DB 892 TAC-----TTCCAGAGACAG-----TGCCATGAATGATGACATACATCAGG 933
QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
DB 934 AAGCAGAACTGGTTTGAGGCTTTGAATACTGATAACTACTAGTAGGCGGCTTAACA 993
QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335
DB 994 AGTGATCTCTTCTTACTTCCAGATGTTACAGGATGTAGTAAATTACTATACTTTTG-- 1050
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 355
DB 1051 -----CGGTGCACGGAACCTGAGGATACAGCTTTTACTATGTG 1086
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
DB 1087 AAATTTTGTACATCCAGAGGTGACAGACCATCCACGTGGGG---AATCAGACTTTT 1143
QY 376 SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTrpPro 393

```

```

DB 1144 ATGTATGGNACTATAGTTCAAAAGTACTTCCGAGAGAGATACAGTACAGTCAAGTAAAGCCA 1203
QY 394 AsnPheIleGluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413
DB 1204 TGGTTAACTGAATCATGATAAT-----TATAAGGTTCTGATCTACAAAT 1248
QY 414 GlyAspAlaAspTyrIle-----CysAsn 421
DB 1249 GSCCAACTGGACATCATGCTGGCAGCTGCCCTGACAGAGCGCTCCCTTGATGGCGATGGAC 1308
QY 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 439
DB 1309 TGGAAAGGATCCAG-----GAATACAAAGAGCGCAGAAAAAAGTT 1350
QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452
DB 1351 TGGAAAGATCTTTAAATCTGACAGTGAAGTGGCTGGTTAC----- 1389
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
DB 1390 -----ATCGGCAAGCGGTGACTTCCATCAGGTAAATTAITCGAGGTGGA 1434
QY 473 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492
DB 1435 GCACATATTTACCGIATGACAGCGCTCTGAGAGCTTTTGACATGATTATCGATTCAIT 1494
QY 493 Phe-----GlyTrpAsp 496
DB 1495 TATGAAAAGGATGGAT 1512
RESULT 11
US-09-904-011-163
; Sequence 163, Application US/09904011
; Publication No. US20030003530A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698

```

; PRIOR FILING DATE: 1999-07-26  
 ; PRIOR APPLICATION NUMBER: US 60/146,222  
 ; PRIOR FILING DATE: 1999-07-28  
 ; PRIOR APPLICATION NUMBER: PCT/US99/20594  
 ; PRIOR FILING DATE: 1999-09-08  
 ; PRIOR APPLICATION NUMBER: PCT/US99/20944  
 ; PRIOR FILING DATE: 1999-09-13  
 ; PRIOR APPLICATION NUMBER: PCT/US99/21090  
 ; PRIOR FILING DATE: 1999-09-15  
 ; PRIOR APPLICATION NUMBER: PCT/US99/21547  
 ; PRIOR FILING DATE: 1999-09-15  
 ; PRIOR APPLICATION NUMBER: PCT/US99/23089  
 ; PRIOR FILING DATE: 1999-10-05  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28214  
 ; PRIOR FILING DATE: 1999-11-29  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28313  
 ; PRIOR FILING DATE: 1999-11-30  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28564  
 ; PRIOR FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28565  
 ; PRIOR FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: PCT/US99/30095  
 ; PRIOR FILING DATE: 1999-12-16  
 ; PRIOR APPLICATION NUMBER: PCT/US99/30911  
 ; PRIOR FILING DATE: 1999-12-20  
 ; PRIOR APPLICATION NUMBER: PCT/US99/30999  
 ; PRIOR FILING DATE: 1999-12-20  
 ; PRIOR APPLICATION NUMBER: PCT/US00/00219  
 ; PRIOR FILING DATE: 2000-01-05  
 ; NUMBER OF SEQ ID NOS: 423  
 ; SEQ ID NO 163  
 ; LENGTH: 2076  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-09-712-338-2-copy\_19\_555 (1-537) x US-09-904-011-163 (1-2076)

Alignment Scores:  
 Pred. No.: 2,99e-32 Length: 2076  
 Score: 359.00 Matches: 130  
 Percent Similarity: 39.51% Conservative: 62  
 Best Local Similarity: 26.75% Mismatches: 178  
 Query Match: 12.44% Indels: 116  
 DB: 11 Gaps: 21

US-09-712-338-2\_copy\_19\_555 (1-537) x US-09-904-011-163 (1-2076)

QY 47 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62  
 DB 295 CCAGGACTGAACATCAAGAGTTATGCCGGCTTCCTCCAGTGAATAAGACTTACACAGC 354  
 QY 63 HisthrPhePheTppPheGluAlaAargHisAsnProGluThrAlaProIleThrLeu 82  
 DB 355 AACCTCTCTCTCTGCTCTCCAGCTCAGATACAGCCAGAGATGCCCCAGTAGTCTC 414  
 QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGlyPro 102  
 DB 415 TGGCTACAGGTGGCCGGAGGTCATCCATGTTGGACTCTTTGTGGAACATGGGCT 474  
 QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTirPAsnGluValSer 122  
 DB 475 TATGTGTACAAAGTAACATGACCTTGGCTGACAGAGACTTCCCGIGGACCACACAGCTC 534  
 QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142  
 DB 535 TCCATGCTTTACATTCACATCCAGTGGCAGAGCTTCAGTTTACTGATGATACCCAC 594  
 QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162  
 DB 595 GGA----- 597  
 QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla 182  
 DB 598 -----TATGCACTCAATGAGGAGGATGATAGCAGGGGATTATATAC 636

QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202  
 DB 637 AGTCCACTAATTCAGTTTTTC-----CAGATATTCTCTGAATAATAAAT 681  
 QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsn 222  
 DB 682 AATCACTTTTATGCTCACTGGGAGTCTTATCAGGAAATATGTGCCACCATTCACAC 741  
 QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242  
 DB 742 CTCAATCCATCCCTCAAC-----CCTGTGACAGAGGTCGAAATCAAC 783  
 QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleLeuAspGluAlaIleGlnAlaProTyrTrp 262  
 DB 784 CTGAACGGAATTTGCTATTGGAGATCGATATCTCATCCGAATCAATATATAGGGGCTAT 843  
 QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 282  
 DB 844 GCAGAAATC-----CTGTACCAATGTGCTTGGATGAGAGCAAAAAG 891  
 QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302  
 DB 892 TAC-----TTCCAGAACAG-----TGCCATGATGATGATAGACACATCAGG 933  
 QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319  
 DB 934 AAGCAGAACTGGTTTGGAGCCCTTTGAAATCTGGATAAACTACTAGATGGCGACTTAACA 993  
 QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335  
 DB 994 AGTGAICCTCTTACTTCCAGAAATGTTACAGAGATGTAGTAATTAATACTTTTGTG--- 1050  
 QY 336 ArgGlyValTyrAspIleArgHisProTyrAspProThrProProSerTyrTyrAsn 355  
 DB 1051 -----CGGTGCAGCGAACCTGAGGATCAGCTTTACTATTGTG 1086  
 QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375  
 DB 1087 AAATTTTGTCTACTCCAGAGGTGAGACAAGCAATCCAGTGGG---AATCAGACTTTT 1143  
 QY 376 SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTrpPro 393  
 DB 1144 AATGATGGAAGTATAGTTGAAAGTACTTGGAGAGATACAGTACAGTACAGTAAAGCCA 1203  
 QY 394 AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413  
 DB 1204 TGGTTAACTGAATCATGAATAAT-----TATAAGTTCTGATCACAAT 1248  
 QY 414 GlyAspAlaAspTyrIle-----CysAsn 421  
 DB 1249 GGCCCACTGGACATCATCGTGGCAGCGCTGACAGAGCGCTCCTTGATGGGCGATGGAC 1308  
 QY 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGln----- 439  
 DB 1309 TGGAAAGGATCCAG-----GAATCAAGAGAGACGAGAAAAGTT 1350  
 QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452  
 DB 1351 TGGAGATCTTTAAATCTGCACAGTGAAGTGGCTGGTTAC----- 1389  
 QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472  
 DB 1390 -----ATCCGGCAAGCGGTGACTTCCATCAGGTAATTAATTCAGGTGGA 1434  
 QY 473 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnPheAsnArgThrIle 492  
 DB 1435 GGACATATTTTACCTATGACCGCTCTGAGAGCTTTTGACATGATTAATCATCATTT 1494  
 QY 493 Phe-----GlyTrpAsp 496  
 DB 1495 TATGGAAGAGATGGCAT 1512

## RESULT 12

US-09-906-742-163  
Sequence 163, Application US/09906742  
Publication No. US20030023054A1  
GENERAL INFORMATION:

APPLICANT: Genentech, Inc.  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, A.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kljavin, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: 10466-14  
CURRENT APPLICATION NUMBER: US/09/906,742  
CURRENT FILING DATE: 2001-07-16  
PRIOR APPLICATION NUMBER: 09/665,350  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: PCT/US99/28214  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30  
PRIOR APPLICATION NUMBER: PCT/US99/28564  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/28565  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30095  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: PCT/US99/30911  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US99/30999  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US00/00219  
PRIOR FILING DATE: 2000-01-05  
NUMBER OF SEQ ID NOS: 423  
SEQ ID NO 163

; LENGTH: 2076

; TYPE: DNA

; ORGANISM: Homo Sapien

US-09-906-742-163

## Alignment Scores:

Pred. No.:	2,99e-32	Length:	2076
Score:	359.00	Matches:	130
Percent Similarity:	39.51%	Conservative:	62
Best Local Similarity:	26.75%	Mismatches:	178
Query Match:	12.44%	Indels:	116
DB:	11	Gaps:	21

US-09-712-338-2\_COPY\_19\_555 (1-537) x US-09-906-742-163 (1-2076)

QY	47	ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer	62
Db	295	CCAGGACTGAACATGAAGAGTTATCCCGGCTTCCTCACCCTGGAATAAGACTTACACAGC	354
QY	63	HISThrPhePheTrpPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu	82
Db	355	AACCTCTCTCTCTGGTCTTCTCCAGCTCAGATACAGCCAGAGATGCCCCAGTAGTTC	414
QY	83	TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro	102
Db	415	IGGCTACAGGTGGCGCGGAGGTTCATCCAGTTTGGAGCTCTTTGGGAACATGGGCT	474
QY	103	CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer	122
Db	475	TATGTTCTCACAATAACATGACCTTGGTGCACAGAGACTTCCCTCGACCACACAGCTC	534
QY	123	AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp	142
Db	535	TCCAIGCTTTACATGACATCCAGTGGCAGCAGGCTTCAGTTTACTGATGATACCCAC	594
QY	143	GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly	162
Db	595	GGA-----	597
QY	163	ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla	182
Db	598	-----TATGCAGTCAATCAGCAGCATGTCAGCAGGATTTATAC	636
QY	183	TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer	202
Db	637	AGTGCACATAATTCAGTTTTC-----CAGATATTTCTGTAATAATAAAAT	681
QY	203	LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsn	222
Db	682	AATCAGTTTATGTCACCTGGGAGCTTATGCAGGGAATATGCCAGCCATTCACAC	741
QY	223	HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn	242
Db	742	CTCATCCATTCCTCAAC-----CCTGTCAGAGAGGTGAAGATCAAC	783
QY	243	PheAsnSerLeuGlyIleIleAsnGlyIleLeuAspGluAlaIleGlnAlaProTyrTyr	262
Db	784	CTCAAGGAAATGCTATTTGGAGATGGATATTCATCCCGAATCAATATAGGGGGCTAT	843
QY	263	ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValIlyAsn	282
Db	844	GCAGAAATC-----CTGTACCAATTTGGCTTGTGGATGAGAGCAAAAAG	891
QY	283	TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys	302
Db	892	IAC-----TTCCAGAAGCAG-----TGCCATCAATGCATAGACACATCAGG	933
QY	303	GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr	319
Db	934	AAGCAGAAGCTGGTTTGAGGCTTTGAAATAACTGATAAATAGTATGCGACTTAACA	993
QY	320	Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly	335

Db 994 AGTGATCCCTCTACTCCAGAAAGTGTACAGAGTGTAGTAATTAATATAACTTTTGG--- 1050  
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProSerTyrIleAsn 355  
Db 1051 -----CGTGCAGGAACCTTGAGGATCAGCTTACTTAAGTG 1086  
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375  
Db 1087 AATATTTTGCATCCAGAGGTGAGACAAGCCATCCACGTGGGG--AATCAGACTTTT 1143  
QY 376 SerAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTrpPro 393  
Db 1144 AATGATCGAAGTATAGTTGAAAGTACTTCCGAGAAAGATACAGTACAGTCAAGTAAAGCCA 1203  
QY 394 AsnPheIleGluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413  
Db 1204 TGGTTAACTGAATCAATGAATAAT-----TATTAAGGTTCTGATCTACAT 1248  
QY 414 GlyAspAlaAspTyrIle-----CysAsn 421  
Db 1249 GGCCAACTGGACATCATCGTGGCGAGCTGCCCTGCACAGAGGCTCTTGATGGCATGGAC 1308  
QY 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGln----- 439  
Db 1309 TGGAAAGGATCCAG-----GAATACAAAGAGCGCAAAAAAAGTT 1350  
QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452  
Db 1351 TGAAGATCTTAAATCTGACATGAGTGAAGTGGCGGTAC----- 1389  
QY 453 ValCltuTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472  
Db 1390 -----ATCCGCAAGCGGGTGGTCTCCATCAGGTAAATTATTCGAGGTGGA 1434  
QY 473 GlyHisCltuValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492  
Db 1435 GCATATATTTTACCTATGACCAAGCTCTGAGAGCTTTTGACATGATTAATCGATTTCAT 1494  
QY 493 Phe-----GlyTrpAsp 496  
Db 1495 TATGGAAGAGATGGGAT 1512

## RESULT 13

US-09-906-838-163  
Sequence 163, Application US/09906838  
Publication No. US20030027143A1

## GENERAL INFORMATION:

APPLICANT: Genentech, Inc.  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, A.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kljavin, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: 10466-14  
CURRENT APPLICATION NUMBER: US/09/906.838  
CURRENT FILING DATE: 2001-07-16  
PRIOR APPLICATION NUMBER: 09/665,350  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: PCT/US99/28214  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30  
PRIOR APPLICATION NUMBER: PCT/US99/28564  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/28565  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30095  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: PCT/US99/30911  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US99/30999  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US00/00219  
PRIOR FILING DATE: 2000-01-05  
NUMBER OF SEQ ID NOS: 423  
SEQ ID NO 163  
LENGTH: 2076  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-09-906-838-163

## Alignment Scores:

Pred. No.: 2,99e-32 Length: 2076  
Score: 359.00 Matches: 130  
Percent Similarity: 39.51% Conservative: 62  
Best Local Similarity: 26.75% Mismatches: 178  
Query Match: 12.44% Indels: 116  
DB: 21 Gaps: 21

US-09-712-338-2\_COPY\_19\_555 (1-537) x US-09-906-838-163 (1-2076)

QY 47 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62  
Db 295 CCAGGACTGAACATGAAGAGTTATCGCGCTTCCTCACCGTGAATAAGACTTACACACAG 354  
QY 63 HisThrPhePheTrpPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82  
Db 355 AACCTCTCTCTCTGGTCTCTCCAGCTCAGATACAGCCAGAGATGCCCGCAGTATTC 414  
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluGlyPro 102  
Db 415 TGGCTACAGGTGGCGCGGAGGTTCAATCCAGTTTGGACTCTTTGTGACATGGGCCI 474  
QY 103 CysHisValAsnSerThrPheAspTyrIleAsnProHisSerTrpAsnGluValSer 122  
Db 475 TATGTGTGCACAAAGTAACATGACCTTGGTGCAGAGAGACTTCCCTGGACCAACAGCTC 534

QY	123	AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp	144
DB	535	TCCATGCTTTTCAATTCAGATCCAGTGGCGACAGGCTTCAGTTTACATGATACCCAC	594
QY	143	GlySerIleLeuAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly	162
DB	595	GGA-----	597
QY	163	ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAla	182
DB	598	-----TAUGCAGTCAATGAGGACGATGTAGCACGGGATTTATAC	636
QY	183	TrpGluIleLeuGlnGlnPheLeuSerGlyLeuProSerLeuAspSerArgValcIuSer	202
DB	637	AGTGCATTAATTCAGTTTTC-----CAGATATTCTCGTAATATAAAAAAT	681
QY	203	LysAspPheSerLeuTTPThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsn	222
DB	682	AATGACITTTAATGACATGGGAGTCTTATGCAGGGAAATATGCCAGCCCATTTGCACAC	744
QY	223	HisPheTyrGluGlnAsnGlnArgIleAlaAsnGlySerValAsnGlyValGlnPheAsn	242
DB	742	CTCATCCATTCCTCAAC-----CCTGTGAGAGAGGTGAAGATCAAC	783
QY	243	PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr	262
DB	784	CTGACAGGAATTCGATTTGGAGATGGATATCTCATCCGAATCAATATATAGGGGGCTAT	843
QY	263	ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn	282
DB	844	CGAAATTC-----CYCTACCAATTCGCTTGTGGATGAGAGCAACAAAAAG	891
QY	283	TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys	302
DB	892	TAC-----ATCCAGAGCAG-----TGCCATGAATCCATAGAACACATCAGG	933
QY	303	GluThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr	319
DB	934	AGCAGACATGGTTTGGGCTTTGGAATACTGCATAAATCTAGATGGCATTAACA	993
QY	320	Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly	335
DB	994	AGTGATCCTTCTACITCCAGAAATTCACAGGATGTAGTAATTACTATACTTTTGTG- -	1050
QY	336	ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn	355
DB	1051	-----CGGTGCACGGAACCTCAGGATCAGCTTTACTATGTG	1086
QY	356	LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln	375
DB	1087	AAATTTTGTGCATCCCGAGAGGTGCAGACAGCCATCCACGTGGGG---AATCAGACATTT	1143
QY	376	SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTyrPro	393
DB	1144	AATGATGGAACTATAGTTGAAAAGTACTTCGGAGAGACATCAGTACAGTCAGTTAAGCCA	1203
QY	394	AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr	413
DB	1204	TGTTAACTGAAATCATCAATAAT-----TATAAGTTCTGCATCTACAAT	1248
QY	414	GlyAspAlaAspTyrIle-----CysAsn	421
DB	1249	GGCCAACTGGACATCATCTGCTGGCGCTGCAGAGCGCTCTCTGTATGGCATGGAC	1308
QY	422	TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln-----	439
DB	1309	TGGAAGGATCCAG-----GAATACAGAAGGCGCAAAAAAAGATT	1350
QY	440	-----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly	452
DB	1351	TGGAAGATCTTTAAATCTGACAGTGAAGTGGCTGGTTAC-----	1389
QY	453	ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla	472

```

Db      1390  -----ATCCGCGCAGCGGTCGACCTTCATCAGGTAATATTCGAGGTGGA 1434
                                     |||||:::|||||
Oy      473  GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492
        ||||| ::|||::: |||||::: :::: ||||| |||
Db      1435  GGACATATTATACCTATGACACGCCCTCTGAGAGCTTTTGACATGATTAATCGATTTCATT 1494
                                     |||||:::|||||
Oy      493  Phe-----GlyTyrAsp 496
        :::: |||||::|
Db      1495  TATGGAAGAGATGGGAT 1512
                                     |||||::|

RESULT 14
US-09-907-613-163
; Sequence 163, Application US/09907613
; Publication No. US20030027145A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Iumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,613
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095

```





```

: APPLICANT: Mather, Jennie P.
: APPLICANT: Pan, James
: APPLICANT: Paoni, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William, I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: FILE REFERENCE: 10466-14
: CURRENT APPLICATION NUMBER: US/09/907,942
: PRIOR FILING DATE: 2002-01-22
: PRIOR APPLICATION NUMBER: PCT/US00/04414
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: US 60/143,048
: PRIOR FILING DATE: 1999-07-07
: PRIOR APPLICATION NUMBER: US 60/145,698
: PRIOR FILING DATE: 1999-07-26
: PRIOR APPLICATION NUMBER: US 60/146,222
: PRIOR FILING DATE: 1999-07-28
: PRIOR APPLICATION NUMBER: PCT/US99/20594
: PRIOR FILING DATE: 1999-09-08
: PRIOR APPLICATION NUMBER: PCT/US99/20944
: PRIOR FILING DATE: 1999-09-13
: PRIOR APPLICATION NUMBER: PCT/US99/21090
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/21547
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/23089
: PRIOR FILING DATE: 1999-10-05
: PRIOR APPLICATION NUMBER: PCT/US99/28214
: PRIOR FILING DATE: 1999-11-29
: PRIOR APPLICATION NUMBER: PCT/US99/28313
: PRIOR FILING DATE: 1999-11-30
: PRIOR APPLICATION NUMBER: PCT/US99/28564
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/28565
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/30095
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: PCT/US99/30911
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US99/30999
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US00/00219
: PRIOR FILING DATE: 2000-01-05
: NUMBER OF SEQ ID NOS: 423
: SEQ ID NO 163
: LENGTH: 2076
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-907-942-163

Alignment Scores:
Pred. No.: 2,99e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.44% Indels: 116
Dbs: 11 Gaps: 21

US-09-712-338-2_COPY_19_555 (1-537) x US-09-907-942-163 (1-2076)

QY 47 ProGly-----ValLysSerThrSerGlyTyr-----ValAspThrSerProGluSer 62
Db 295 CCAGGACTGAACATGAAGAGTATGCCGCTTCCCTCAGCGTGAATAAGACTTTACAACAGC 354
QY 63 HisThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
Db 355 AACCTCTCTCTCTGTTCTTCCAGCTCAGATACAGACAGAGATGCCCGCTGAGTTC 414
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGlyPro 102

```

```

Db 415 TGGCTACAGGIGGGCCGGAGGTCATCCAGTCTTGGAGCTCTTGTGGAACATGGGCT 474
QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 122
Db 475 TATGTTGTCACAAGTAACATGACCTCCGTCACAGAGACTTCCCTGGACACACAGGTC 534
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerThrValAsp 142
Db 535 TCCATGCTTTACATGATCCAGTGGCAGAGCTTCACTTCTACTGATGATACCCAC 594
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
Db 595 GGA----- 597
QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla 182
Db 598 -----TATGCAGTCAATGAGCAGATGATGACGGGATTATAC 636
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
Db 637 AGTGCACATAATTCAGTTTTC-----CAGATAITTCCTGAATAATAAAAT 681
QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsn 222
Db 682 AATGACTTTTATGTCACCTGGGAGCTTATGCGAGGAAATATGCGCAGCATTCACAC 741
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
Db 742 CTCATCCATCCCTCAC-----CCTGTGAGAGAGTGAAGATCAAC 783
QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 262
Db 784 CTGACGGAATTCGATTGAGATGGATATCTGATCCGGAATCAATATAGGGGGCTAT 843
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 282
Db 844 GCAGAAATC-----CTGTACCAAAATGGCTTGTGGATGAGACGAAAAAAG 891
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
Db 892 TAC-----TTCAGAGAGCAG-----TGCCATGAATGCATAGAACACATCAGG 933
QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
Db 934 AAGCAGAACTGGTTTGAGGCTTTGAAATACTGGATAAACTACTAGATGCGACTTAACA 993
QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335
Db 994 AGTATCCCTTCTTACTTCCAGAAATGTACAGAGTAGTAATTAATACTAATACTTTTG 1050
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 355
Db 1051 -----CGGTGACGGAACCTCAGGATCAGCTTACTATGTC 1086
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
Db 1087 AAAATTTTGICACTCCAGAGGTGAGACAAAGCATCCACGTGGG---AATCAGACTTTT 1143
QY 376 SerAsnAsnAspValTyr-----TyrAlaPheGlnGluThrGlyAspPheValIrrPro 393
Db 1144 AATGATGGAACTAATAGTGAAGACTTTCGCGAGAGATACAGTACAGTACAGTAAAGCCA 1203
QY 394 AsnPheIleGluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413
Db 1204 TGGTTAAGTGAATCATGATATAT-----TATAGGTCTTGATCTACAAT 1248
QY 414 GlyAspAlaAspTyrIle-----CysAsn 421
Db 1249 GCGCAACTGGACATCAICGTGGCAGCTGCCCTGCAGAGCGCTCCTTGATGGCGATGAC 1308
QY 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 439

```

```
Db 1309 TGGAAAGGATCCAG-----GAATACAGAGAGGCGAGAAAAAAGTT 1350
QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452
Db 1351 TGGAGAGATCTTTAAATCTGACAGTGAAGTGGCTGTTAC----- 1389
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
Db 1390 -----ATCGGCAAGGGGTGACTTCCATCAGTAAATTATTCGAGGTGA 1434
QY 473 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnPheAsnArgThrIle 492
Db 1435 GGACATATTTTACCTATGACAGGCTCTGAGAGCTTTTTCACATGATTATTCGATTTCATT 1494
QY 493 Phe-----GlyTyrAsp 496
Db 1495 TATGGAAGAGGATGGAT 1512
```

Search completed: September 17, 2003, 01:54:21  
Job time : 466 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 16, 2003, 22:53:23 ; Search time 3376 Seconds  
(without alignments)  
3865.967 Million cell updates/sec

Title: US-09-712-338-2\_COPY\_19\_555

Perfect score: 2887  
Sequence: 1 LPGSTPASVGRRLPKNPIG.....HTQSSVPLPTATSMSSVGM 537

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool.D/US09712338/runat\_16092003\_144914\_19142/app\_query.fasta\_1.711  
-DB=EST -QPM=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09712338 @CGN\_1.1\_3549 @runat\_16092003\_144914\_19142 -NCPU=3  
-NO\_MMAL -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: em\_estba:\*  
2: em\_estum:\*  
3: em\_estin:\*  
4: em\_estm:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_pqg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	512	17.7	588	9	AW672518	AW672518 LG1_360_F
2	409	14.2	1049	29	CNS05XWN	AL420285 T3 end of
3	405.5	14.0	648	10	BE777028	BE777028 MY-23-G-0
4	405.5	14.0	653	14	CD258740	CD258740 pLMY023ag
c 5	403.5	14.0	907	14	CD456151	CD456151 Fg03_10q0
c 6	385.5	13.4	1002	29	CNS06ZOK	AL422586 T3 end of
7	384	13.3	1054	29	CNS06NN7	AL406985 T7 end of
8	375.5	13.0	986	29	CNS06Q80	AL410350 T7 end of
9	354	12.3	1005	29	CNS06EAK	AL394866 T7 end of
10	353	12.2	3268	11	AK032197	AK032197 Mus muscu
11	352.5	12.2	963	29	CNS06ZSE	AL421896 T3 end of
12	349	12.1	2493	11	AK019861	AK019861 Mus muscu
13	345	12.0	721	14	CA747625	CA747625 GAP13D07
14	333	11.5	1581	11	AF217508	AF217508 Homo sapi
15	332.5	11.5	712	14	CD045123	CD045123 pSHB017xN
16	332.5	11.5	729	13	BQ110016	BQ110016 VD0102C06
17	327.5	11.3	685	13	BW251033	BW251033 BW251033
18	327.5	11.3	700	13	BW246209	BW246209 BW246209
19	326	11.3	1492	11	AK017087	AK017087 Mus muscu
20	324.5	11.2	876	14	CD384438	CD384438 PTM009611
21	322	11.2	652	9	AV991679	AV991679 AV991679
22	321	11.1	879	29	CNS0750D	AL429491 cLore BA0
23	320.5	11.1	675	13	BW274166	BW274166 BW274166
24	320.5	11.1	675	13	BW302983	BW302983 BW302983
25	320.5	11.1	680	13	BW208298	BW208298 BW208298
26	320.5	11.1	684	13	BW248944	BW248944 BW248944
27	320.5	11.1	685	13	BW194502	BW194502 BW194502
28	320.5	11.1	686	13	BW260398	BW260398 BW260398
29	320.5	11.1	689	13	BW046109	BW046109 BW046109
30	320.5	11.1	692	13	BW194721	BW194721 BW194721
31	320.5	11.1	699	13	BW035379	BW035379 BW035379
32	319.5	11.1	566	13	BW276795	BW276795 BW276795
33	319.5	11.1	656	13	BW276796	BW276796 BW276796
34	319.5	11.1	702	13	BW212011	BW212011 BW212011
35	319.5	11.1	762	13	BQ165650	BQ165650 EST611519
36	318	11.0	676	14	CD044351	CD044351 pSHB014XA
37	318	11.0	700	12	B1749802	B1749802 Fg02_05d0
38	316.5	11.0	682	13	BW255378	BW255378 BW255378
39	313	10.8	485	12	B1200653	B1200653 Oia061fs.f
40	311.5	10.8	638	13	BW282449	BW282449 BW282449
41	311	10.8	708	14	CD213567	CD213567 HSL_4.L.F1
42	309.5	10.7	656	13	BW292110	BW292110 BW292110
43	308	10.7	804	14	CB644508	CB644508 OSJNE050M
44	307.5	10.7	580	13	BW197413	BW197413 BW197413
45	307	10.6	637	14	CA352124	CA352124 623325 NC

ALIGNMENTS

RESULT 1  
AW672518  
LOCUS LG1\_360\_F03.b1\_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA  
DEFINITION sequence.  
ACCESSION AW672518  
VERSION AW672518.1 GI:7536439  
KEYWORDS EST.  
SOURCE Sorghum bicolor (sorghum)  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 588)  
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.  
An EST database from Sorghum: Light-grown seedlings  
Unpublished  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.  
Seq primer: JEN REV  
High quality sequence stop: 580  
POLYA-No.  
Location/Qualifiers  
1..588  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/db\_xref="taxon:4558"  
/clone\_lib="Light Grown 1 (LGI)"  
/notes="Organ: 10- to 14-day-old light-grown (greenhouse)  
seedlings; Vector: lambda Zap; Site\_1: XhoI; Site\_2: EcoRI  
; The library was made from poly-A RNA in the cloning  
vector lambda ZAP II. Clones to be sequenced were  
prepared by mass excision."  
BASE COUNT 145 a 151 c 152 g 140 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,09e-46 Length: 588  
Score: 512.00 Matches: 96  
Percent Similarity: 71.15% Conservative: 15  
Best Local Similarity: 61.54% Mismatches: 41  
Query Match: 17.73% Indels: 4  
DB: Gaps: 2  
US-09-712-338-2\_COPY\_19\_555 (1-537) x AW672518 (1-588)  
Qy 14 LeuProLysAsnProThrGlyValLysThrLeuThrAlaAsnAsnValThrIleArg 33  
Db 128 CTCGCAAGGAGTACATGACTACAGACGGCTACTGCACCAACAATGTGCATCTCGA 187  
Qy 34 TyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyrSer 53  
Db 188 TACAGAACCT-----GGTATCTGTGAGACGCGCTGTGTGTACAGACTACTCA 238  
Qy 54 GlyTyrValAspThrSerProGluSerHisThrPhePheThrPheGluAlaArgHis 73  
Db 239 GGATACGGTGGATCTCGCACCAATGTCACAGCTTCTTCTGTTCTTCGAGCTAGGAAC 298  
Qy 74 AsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeu 93  
Db 299 AATCCGCATCTGATCGCTTCACTTTGTTGTTGATGGCGGTCCCGGATCATCGTCTG 358  
Qy 94 IleGlyLeuPheGluGluLeuGlyProCysHisValAsnSerThrPheAspTyrIle 113  
Db 359 ATTGGGCTCTTTGAAGAAAGGTCATGCAATGATGATGATATCTCACGGCTGAGTAC 418  
Qy 114 AsnProHisSerTrpAsnGluValSerAsnLeuPheLeuSerGlnProLeuGlyVal 133  
Db 419 AACCCCTACAGCTGGAACAATGCTCTCAACATGCTATATTTCCAGCCGCTCGGCAGC 478  
Qy 134 GlyPheSerTyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGlu 153  
Db 479 GGATTCAGCTACCGAGAACAGCGCTCGCGCTTGAATTCAGTCCAGCAGCAATTCGCT 538  
Qy 154 ---AsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAsp 168  
Db 539 TACAACCTCGAGTGAATACCGGCTGATGGCGGTACCGGCTCTTGAT 586

RESULT 2  
CNS06XWN 1049 bp DNA linear GSS 06-JUL-2001  
LOCUS T3 end of clone AY0AA004D06 of library AY0AA from strain CBS 6340  
DEFINITION of Kluyveromyces thermotolerans, genomic survey sequence.  
ACCESSION AL420285  
VERSION AL420285.1 GI:12203469  
KEYWORDS GSS.  
SOURCE Kluyveromyces thermotolerans  
ORGANISM Kluyveromyces thermotolerans  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
REFERENCE 1 (bases 1 to 1049)  
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
Bollotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,  
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,  
Malpertuy,A., Neugeglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
Saurin,W., Tekaja,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
Wincker,P. and Weissenbach,J.  
Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies  
FEBS Lett. 487 (1), 3-12 (2000)  
JOURNAL 20584711  
MEDLINE 11152876  
PUBMED 11152876  
REFERENCE 2 (bases 1 to 1049)  
AUTHORS Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P.  
and Dujon,B.  
Genomic exploration of the hemiascomycetous yeasts: 10.  
Kluyveromyces thermotolerans  
FEBS Lett. 487 (1), 61-65 (2000)  
JOURNAL 20584720  
MEDLINE 11152885  
PUBMED 11152885  
REFERENCE 3 (bases 1 to 1049)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
COMMENT This GSS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces marxianus var. marxianus, Pichia  
lactis var. lactis, Kluyveromyces thermotolerans, Kluyveromyces  
angustis, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.  
FEATURES  
Location/Qualifiers  
1..1049  
/organism="Kluyveromyces thermotolerans"  
/mol\_type="genomic DNA"  
/strain="CBS 6340"  
/db\_xref="taxon:4916"  
/clone="AY0AA004D06"  
/clone\_lib="AY0AA"  
/note="end : T3"  
misc\_feature <82..>1020  
/note="similar to Saccharomyces cerevisiae ORF YBR139W [  
strong similarity to carboxypeptidase ]"  
/evidence=not\_experimental  
misc\_feature <172..>1017  
/note="similar to Saccharomyces cerevisiae ORF YMR297W [  
PRC1 ; carboxypeptidase Y, serine-type protease ]"  
/evidence=not\_experimental  
BASE COUNT 262 a 253 c 252 g 280 t 2 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.38e-34 Length: 1049  
Score: 409.00 Matches: 114  
Percent Similarity: 43.90% Conservative: 37  
Best Local Similarity: 33.14% Mismatches: 137

Query Match: 14.17% Indels: 56  
 DB: 29 Gaps: 10  
 US-09-712-338-2\_COPY\_19\_555 (1-537) x CNS06XWN (1-1049)

QY 1 LeuProGlySerThrProAlaSerValGly-----ArgArgGlnLeuProLysAsnPro 18  
 DB STCCCAAAATACCGTCAGCTATCAAAATGTCGTAAGTCTGAGAGTTCGTC 168  
 QY 19 ThrGlyValLysThrLeuThrAlaAsnValThrIleArgTyrLysGluProGly 38  
 DB 169 ACGATAAGCTGAGCTTCATGACACCTACATGAGGGTAAGAGAGAGTGCACCTTCG 228  
 QY 39 AlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThr 58  
 DB 229 AAGCTGGGTGC-----GATTCIGTCAAGCAGTACTCGGGTATCTGGACAC 276  
 QY 59 SerProGluSerHisThrPhePheThrPheGluAlaArgHisAsnProGluThrAla 78  
 DB 277 GAGGATTCACAACTCTTTTACTGGCATTTGAATCCAGAAACGCCCACTCAATGAC 336  
 QY 79 ProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGlu 98  
 DB 337 CCCGTATTCTATGGCTGAACGGTGGTCCAGGGTCTCTTTCACGGGTTATTTTC 396  
 QY 99 GluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrp 118  
 DB 397 GAACCTGGCCCTCTTCGTCGCGCTGAGCTAAGCCGTCGCAATCCGACTCCGTCG 456  
 QY 119 AsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSer 138  
 DB 457 AATAACATGCTACTGTCATATCTTGAGCAGCGCTTAGCGTTGGTTTCTCTATG 516  
 QY 139 AspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAla 158  
 DB 517 GAC----- 519  
 QY 159 GlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAla 178  
 DB 520 -----GAGAGTGTGCTCCACACAAACGACCA 546  
 QY 179 AlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSer 198  
 DB 547 GGAAGATGTTGTCATATCTTGAGCTGTTTTCGAAGAGTTTCCCA----- 594  
 QY 199 ArgValGlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyPro 218  
 DB 595 CAGTTTAGATCTACGACTTCATATGTCAGGTGAATCATACGCTGGCCATTACATCCCA 654  
 QY 219 AlaPhePheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGly 238  
 DB 655 -----GAGATGCCCAATCAATCGCTGCTGTCATGATCTGAC 693  
 QY 239 ValGlnLeuAsnPheAsnSerLeuGlyIleLeuAsnGlyIleLeuAspGluAlaIleGln 258  
 DB 694 ABAACCTCAATCTACCTCCATGATGGCAACGGGATAACAGATTCTTTAGTCCAG 753  
 QY 259 AlaProTyrThrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGlu 278  
 DB 754 TATGACTACTATGAGCCCATGGCTGGCAGGGGT---GGTTACAAGCGCGT---ATA 807  
 QY 279 ThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIle 298  
 DB 808 ACTGAGSAGAAATGCGCAAAATGAGAAACCAATGCTGCG---TGCAGAGCTTTGAC 864  
 QY 299 SerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAla 318  
 DB 865 AATGCACTGACTACCTCTAGTTCCAGC-----TCGCTTGTATTGACGCT 909  
 QY 319 ThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyVal 338  
 DB 910 GCGCCATCTGCT---GAGAAATATGGCCATGAGTGCCTTACACTAAACAGGCGCTGAAGCT 966

QY 339 TyrAspIleArg 342  
 DB 967 TATGACATACGC 978

RESULT 3  
 BE777028  
 LOCUS MY-23-G-01 PinfestansMY 648 bp mRNA linear EST 20-SEP-2000  
 DEFINITION Phytophthora infestans cDNA, mRNA sequence.  
 BE777028  
 VERSION BE777028.1 GI:10230683  
 KEYWORDS EST.  
 SOURCE Phytophthora infestans (potato late blight agent)  
 ORGANISM Phytophthora infestans  
 Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;  
 Phytophthora.  
 1 (bases 1 to 648)  
 Kamoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.  
 Initial assessment of gene diversity for the oomycete pathogen  
 Phytophthora infestans based on expressed sequences  
 Fungal Genet. Biol. 28 (2), 94-106 (1999)  
 20056376  
 MEDLINE 10587472  
 COMMENT Contact: Govers F  
 Laboratory of Phytopathology  
 Wageningen University  
 Binnenhaven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands  
 Tel: 31 317 483 138  
 Fax: 31 317 483 412  
 Email: Francine.Govers@medew.fyto.wau.nl.

FEATURES  
 source  
 1..648  
 /organism="Phytophthora infestans"  
 /mol\_type="mRNA"  
 /strain="DDR7602, Al mating type"  
 /db\_xref="taxon:4787"  
 /dev\_stage="4-week old vegetative, non-sporulating  
 mycelium in synthetic medium"  
 /lab\_host="E. coli, strain DH5-alpha"  
 /clone\_lib="PinfestansMY"  
 /note="Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; Total  
 RNA was isolated from mycelium of P. infestans DDR7602  
 cultured for 4 weeks in synthetic medium. EST clones were  
 named by their position in the microtiter plate, preceded  
 by the prefix MY (for mycelial) and the successive number  
 of the microtiter plate (e.g. MY-06-A-04)."

BASE COUNT 161 a 174 c 168 g 145 t  
 ORIGIN

Alignment Scores:  
 Pred No.: 1.75e-34 Length: 648  
 Score: 405.50 Matches: 86  
 Percent Similarity: 50.63% Conservative: 35  
 Best Local Similarity: 35.98% Mismatches: 77  
 Query Match: 14.05% Indels: 41  
 DB: 10 Gaps: 6

US-09-712-338-2\_COPY\_19\_555 (1-537) x BE777028 (1-648)

QY 40 GluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyr-----ValAsp 57  
 DB 48 GAGGATTTTGTGACTGACG-----AAGCAGTTGAGTGGCTACTTCAAGATAACA 98  
 QY 58 ThrSerProGluSerHisThrPhePheThrPheGluAlaArgHisAsnProGluThr 77  
 DB 99 GGCTCCAAGTCGAAGAACTACTTTTACTGTCTTCGATCACGCGCAGTCCCTCAGCC 158  
 QY 78 AlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPhe 97  
 DB 159 GATCCCGCTGATCATCTGGCTTACTGGCGGTCCTGGATGCACTTCTGGCTTACTG 218  
 QY 98 GluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSer 117  
 DB 219 CAGAAACAGCCCGCTGCTCCGTGAAGCAGGACTGTCTACICAGAAAGAAATCGTATCG 278

Qy	118	TrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyr	137		
Db	279	TGGAAACGAGCGTGCCATGTCAGTGCATGACACGAGCCGCTGGGTGTGGCTTCAGCTAC	338		
Qy	138	SerAspThrValAspGlySerIleAsnProValThrGlyValGluAsnSerSerPhe	157		
Db	339	GGAGAC-----	344		
Qy	158	AlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeu	177		
Db	345	-----AGCGGTGAATACGACACGTCAGACGAGGAGGCGCGACGAC-----	386		
Qy	178	AlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAsp	197		
Db	387	-----ATGTTCCACTTCTCGAAGAATCTTCAAGGCTTGCCA-----	425		
Qy	198	SerArgValGlnSerIleAspPheSerLeuThrPheGluSerTyrGlyGlyHisTyrGly	217		
Db	426	---GAGTATCAAAAACCTCCCTTACGCTCTTCGAGAGAGATTACGCGGCTACGCTACG	482		
Qy	218	ProAlaPhePheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsn	237		
Db	483	CCCGGATCGCTCACAGAACTTTTACTTGGCAACACGCAA-----	530		
Qy	238	GlyValGlnLeuAsnPheAsnSerLeuGlyIleLeuAsnGlyIleIleAspGluAlaIle	257		
Db	531	CCTGTGGAATAATTTGAAGGCTTTTGAATTTGGAACTGGCAACGCA-----	590		
Qy	258	GlnAlaProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLeuLysAlaVal	276		
Db	591	CAGTACAACTACTACCCGACATGCGCTTACCAACAATACATACGCGCTCAGGCTGTG	647		
RESULT 4	CD258740	653 bp	mRNA	linear	EST 23-MAY-2003
LOCUS	piMY023aG01r_213782	Wageningen/U.Maryland Phytophthora infestans			
DEFINITION	EST Project Phytophthora infestans cDNA clone IMY023G01 5, mRNA				
ACCESSION	CD258740				
VERSION	CD258740.1	GI:31046567			
KEYWORDS	Phytophthora infestans (potato late blight agent)				
SOURCE	Phytophthora infestans				
ORGANISM	Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora.				
REFERENCE	1 (bases 1 to 653)				
AUTHORS	Kamoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.				
TITLE	Initial assessment of gene diversity for the oomycete pathogen Phytophthora infestans based on expressed sequences				
JOURNAL	Fungal Genet. Biol. 28 (2), 94-106 (1999)				
MEDLINE	20056376				
PUBMED	10587472				
COMMENT	Contact: Tyler B Tyler lab VBI 1880 Pratt Dr., Blacksburg, VA 24061, USA Tel: 540-231-7318 Email: bmtyle@vt.edu PCR Primers FORWARD: M13 reverse 17mer at 5' end BACKWARD: M13 reverse 17mer at 5' end Plate: 023 row: G column: 01 Seq primer: M13 reverse 17mer at 5' end High quality sequence stop: 653. Location/Qualifiers 1..653 /organism="Phytophthora infestans" /mol_type="mRNA" /db_xref="taxon:4787" /clone="IMY023G01" /tissue_type="mycelium" /cell_line="DDR7602"				
FEATURES	source				
	CD456151/c				
	LOCUS				
	DEFINITION				
	Fg03_10g07_R				
	ion Gibberella zeae cDNA clone Fg03_10g07, mRNA sequence.				
	CD456151				
	ACCESSION				
	CD456151.1	GI:31370891			
	907 bp	mRNA	linear		EST 03-JUN-2003

KEYWORDS EST.  
SOURCE Gibberella zeae  
ORGANISM Gibberella zeae  
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
AUTHORS 1 (bases 1 to 907)  
Harris,L.J., Allard,S., Rochleau,H., Chapados,J., Couroux,P.,  
Hattori,J., Masotti,M., Ouellet,T., Robert,L.S., Singh,J.A., Sprott  
D. and Tinker,N.A.  
TITLE A cDNA library prepared from Fusarium graminearum grown in liquid  
culture under trichothecene-producing conditions  
JOURNAL Unpublished  
COMMENT Contact: Harris, Linda J.  
Eastern Cereal and Oilseed Research Centre  
Agriculture and Agri-food Canada  
Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,  
CANADA  
Tel: (613) 759-1314  
Fax: (613) 759-6566  
Email: harrislj@em.agr.ca.  
FEATURES  
source Location/Qualifiers  
1. 907  
/organism="Gibberella zeae"  
/mol\_type="mRNA"  
/strain="DAOM 180378"  
/db\_xref="taxon:5518"  
/clone="Fg03\_10g07"  
/tissue\_type="Mycelium"  
/dev\_stage="Asexual"  
/lab\_host="E. coli SURE2"  
/clone\_lib="Fg03\_AAFc\_EOORC\_Fusarium\_graminearum\_mycelium\_  
trichothecene\_production"  
/note="Vector: pGEM-T easy; Site.1: EcoRI; Fusarium  
graminearum grown in liquid culture to optimize  
trichothecene production; final deoxyribonucleic  
concentration in culture determined by CD-ELISA to be 165  
micrograms/ml. cDNA was prepared using a modified SSH  
protocol to promote normalization whereby cDNA was not  
digested prior to hybridization and no subtraction was  
attempted"  
BASE COUNT 243 a 192 c 225 g 242 t 5 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 4,64e-34 Length: 907  
Score: 403.50 Matches: 101  
Percent Similarity: 51.30% Conservative: 37  
Best Local Similarity: 37.55% Mismatches: 120  
Query Match: 13.98% Indels: 12  
DB: 14 Gaps: 5  
US-09-712-338-2\_COPY\_19\_555 (1-537) x CD456151 (1-907)  
QY 224 PheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPhe 243  
DB 906 TTGAACGACAGAACTGCTGATATGATGCTGATACCAATCAAGGGCGAGAGATCAAGSTC 847  
QY 244 AsnSerLeuGlyIleIleAsnGlyIleAspGluAlaIleGlnAlaProTyrTyrPro 263  
DB 846 GTAGCTCTGGCATCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 787  
QY 264 GluPheAlaValAsnAsnThrTyrGlyIleAspGluAlaValAsnGluThrValTyrAsnIyr 283  
DB 786 GAATAGCGACCAATAACACCTACAAA---AAGTGTATCAGCTCACT-----CAGTAC 736  
QY 284 MetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuSerThrCysLysGln 303  
DB 735 AGCAAGTCACTAAACACCTACATCAAAAGTGTGCTGCTTTCGCGAAGTCCAGGCC 676  
QY 304 ThrAsnArgThrAlaLeuAlaAspTyrAlaLeuGlyCysAlaGluAlaThrAsnMetCysArg 323  
DB 675 TCACAGGCAACGACGCT-GCT-----TGTGGCAATGCCGACGATGCTGCAGT 629

QY 324 AspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHis 343  
DB 628 CAGGCTATGAACGGCCCTTGGAGAAATACAGACAGACTTIGAGCTTATGATATCGGGCC 569  
QY 344 ProTyrAspProThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerVal 363  
DB 568 CGAGTAACGACCCCTTCCCTCACTACGACTTACGACTTACGACTTACGACTTACGACTT 509  
QY 364 MetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspValTyrAla 383  
DB 508 ATGAAGCTATTGGCGCGCAGTCAACATACGGCGAATCTCCGAGCGGCTTATGACAG 449  
QY 384 PheGlnGlnThrGlyAspPheValTyrProAsnPheIleGluAspLeuGluIleLeu 403  
DB 448 TTTATCAACAGTGGCGAT---CGGGGACGATCAATCTTGTCCACATTTGCCAAGTCATT 392  
QY 404 AlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPhe 423  
DB 391 GACTCCAAGATCAAGTATTGATCTGGCTGGGATGGCGGACTGGATCTGCACTGGATG 332  
QY 424 GlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAla 443  
DB 331 GGTATTACCGC-----GCCTCAACTCTATTGCTGCCAGCTGCTCTCTCAGCT 281  
QY 444 GlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsn 463  
DB 280 CCTCTCAATCCTTACTGTTGATGGAACCTAAGTACGAGAAATTAACACATCTGGAAT 221  
QY 464 PheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAla 483  
DB 220 ITGAGCTGGCTGGCGAGTTATGCTGCTCATGAGTCCCGCTTATCAGCCCTCAGCT 161  
QY 484 SerLeuGlnLeuPheAsnArgThrIle 492  
DB 160 CATTTGGCTGCTTTTTCGCCACACTT 134  
CNS06ZOK 1002 bp DNA linear GSS 06-JUL-2001  
T3 end of clone AZ0AA003B09 of library AZ0AA from strain CBS 712 of  
Kluyveromyces marxianus, genomic survey sequence.  
AL422586  
AL422586.1 GI:12205780  
GSS.  
Kluyveromyces marxianus  
Kluyveromyces marxianus  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
1 (bases 1 to 1002)  
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,  
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,  
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
Wincker,P. and Weissenbach,J.  
Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies  
FEBS Lett. 487 (1), 3-12 (2000)  
20584711  
PUBMED  
11152876  
REFERENCE 2 (bases 1 to 1002)  
Llorente,B., Malpertuy,A., Blandin,G., Artiguenave,F., Wincker,P.  
and Dujon,B.  
Genomic exploration of the hemiascomycetous yeasts: 12.  
Kluyveromyces marxianus var. marxianus  
FEBS Lett. 487 (1), 71-75 (2000)  
20584722  
PUBMED  
11152887  
REFERENCE 3 (bases 1 to 1002)  
Genoscope.  
Direct Submission  
TITLE  
JOURNAL  
AUTHORS  
PUBMED  
JOURNAL  
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :

seqref@genoscope.cns.fr - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

## FEATURES

Location/Qualifiers  
 1..1002  
 /organism="Kluyveromyces marxianus"  
 /mol\_type="genomic DNA"  
 /strain="CBS 712"  
 /variety="marxianus"  
 /db\_xref="taxon:4911"  
 /clone\_lib="AZ00A003B09"  
 /clone\_lib="AZ00A"  
 /note="end : T3"  
 complement(<8..>994)  
 /note="similar to Saccharomyces cerevisiae ORF YMR297w [PRC1 ; carboxypeptidase Y, serine-type protease ]"  
 /evidence="not experimental"  
 /note="similar to Saccharomyces cerevisiae ORF YBR139w [complement(<11..>778)  
 strong similarity to carboxypeptidase ]"  
 /evidence="not experimental"  
 267 a 259 c 240 g 235 t 1 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 5.38e-32 Length: 1002  
 Score: 385.50 Matches: 102  
 Percent Similarity: 42.59% Conservative: 36  
 Best Local Similarity: 31.48% Mismatches: 123  
 Query Match: 13.35% Indels: 63  
 Gaps: 11

US-09-712-338-2\_COPY\_19\_555 (1-537) x CNS06ZOK (1-1002)

Qy 18 ProThrGlyValLysThrLeuThrAlaAsnValThrIle----- 32  
 Db 826 CCAGAGAGCTTTGTGTCTACGGCTAGCGTCAACAGCAGCAATGTGCGGAAGTACCAGTIG 767  
 Qy 33 -----ArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyVal 49  
 Db 766 CGTGTGAAGAAGATCAAGGACCCCTAAGGTGCTTGGCGTG-----GACCCCGACGTG 716  
 Qy 50 LysSerTyrSerGlyTyrValAspThrSerProGlu---SerHisThrPhePheThrPhe 68  
 Db 715 AAGCAATACCTCGGTACTTGGAGCTTGAACAGGACAGCAGCACTTCTTCTACTGGTTC 656  
 Qy 69 PheGluAlaArgHisAsnProGluThrAlaProThrLeuThrLeuAsnGlyGlyPro 88  
 Db 655 TTTGAGTCCAGAAACGACCCCAAGACACCTGTTATCTGTGTGTCAACGGTGGGCCA 596  
 Qy 89 GlySerAspSerLeuLeuGlyLeuPheGluGluLeuGlyProCysHisValAsnSerThr 108  
 Db 595 GGGGTTCCTCTTTGACTGCTTTGTTCTTCGAAATGGGTTCCTTCATTCCTGGCGAAGAG 536  
 Qy 109 PheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuPheLeuSer 128  
 Db 535 GTGAGCCAAATTACACCCACATCTGGACACCAACGCTTCGGTTATCTTCTTGGAC 476  
 Qy 129 GlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnProVal 148  
 Db 475 CAGCCAGTCAAGCTGGTTACTTCTTACTCTTCAATCA---GAAGGTGCTCAACACTGTT 419  
 Qy 149 ThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAsp 168  
 Db 418 GCT----- 416

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

169 AlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluLeuLeuGlnGly 188  
 415 -----GCAGTAAGGACGCTGTATGCGTCTTCTTCAATG 383  
 189 PheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPheSerLeuTrp 208  
 382 TTCTTCCAGCAATTCCTCCGAGTAGCCTCT-----GCCCAAGATTTCCACATTGCT 332  
 209 ThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsnHisPhe-----TyrGlu 226  
 331 GGTGAATCTTATGCGGGCCACTATCCTCTGTGTTGCTACAGAGATTTTTCGCCCA 272  
 227 GlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeu 246  
 271 CAAGAGGAGCT-----TCTTTCAACTTGACCTCTGTT 239  
 247 GlyIleLeuAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAla 266  
 238 TTGATCGTAAACGGTTTGACCGCCCTTGACCGAGGAGCAATGTGAAGCATG----- 179  
 267 ValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValIyrAsnTyrMetLysPhe 286  
 178 TGTGTCAGGTGTGAGCCCATCTGTGTGACGAGGAGCAATGTGAAGCATG----- 125  
 287 AlaAsnGlnMetProAsnGlyCysGlnAspLeuLeuSerThrCysLysGlnThrAsnArg 306  
 124 TTGGAAGCGTACCCAGA--TGCTAAACTGATTCATCTTGTACGAA----- 77  
 307 ThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnVal 326  
 76 -----TACGAATCTGTTGTGTCATGTGTGCGAGCTCCCTACTACTGTAAACAGCTCAA 23  
 327 GluGlyProTyr 330  
 22 ATGGTCCATAC 11  
 RESULT 7  
 CNS06NN7  
 LOCUS  
 DEFINITION  
 T7 end of clone AU0AA015E08 of library AU0AA from strain CBS 3082  
 of Saccharomyces kluyveri, genomic survey sequence.  
 AL0406985.1 GI:12171743  
 VERSION  
 AL0406985.1 GI:12171743  
 KEYWORDS  
 GSS.  
 SOURCE  
 Saccharomyces kluyveri  
 ORGANISM  
 Saccharomyces kluyveri  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 REFERENCE  
 1 (bases 1 to 1054)  
 Souciet,J.L., Algle,M., Artiguenave,F., Blandin,G.,  
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,  
 de-Montigny,J., DuJon,B., Durrens,P., Lepingle,A., Llorente,B.,  
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
 Wincker,P. and Weissenbach,J.  
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
 FEBS Lett. 487 (1), 3-12 (2000)  
 20584719  
 11152876  
 11152876  
 2 (bases 1 to 1054)  
 Neuvéglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,  
 Gaillardin,C. and Casaregola,S.  
 Genomic exploration of the hemiascomycetous yeasts: 9.  
 Saccharomyces kluyveri  
 FEBS Lett. 487 (1), 56-60 (2000)  
 20584719  
 11152884  
 3 (bases 1 to 1054)  
 Genoscope.  
 Direct Submission  
 Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,  
 JOURNAL











QY 49 ValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePheTirpPhe 68  
 Db 169 TTCGGCAATACTCCGGCTACTCAGAGCATCCAGACTCCACAGCATTCACACTACGTGTTT 228  
 QY 69 PheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyGlyPro 88  
 Db 229 GTGAGTGCAGAACAGAGCCAAAGACAGACCCCGTGGTGGCTTGGCTTAAGGGGGTCC 288  
 QY 89 GlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsnSer--- 107  
 Db 289 GGTTCGACGCTCGCTCGATGGCTGCTACAGAGCAGCGCCCTTCTCATCCAGCCAGAT 348  
 QY 108 -----ThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu 125  
 Db 349 GGTTCACCCCTGGAG-----TACACCCCTATCTGTGAACCTGATGCAACGTGCTG 402  
 QY 126 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 145  
 Db 403 TATATCAGTCCCGCAGCTGGGTGGCTTCCTTACTCGAT----- 444  
 QY 146 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 165  
 Db 444 ----- 444  
 QY 166 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaIrpGluIle 185  
 Db 445 -----GACAGATGTCAGTGACCAATCAGACAGAGTGGCGGAGACAATATGAGCC 498  
 QY 186 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 205  
 Db 499 CTTAAAGACTTCTTCGGCTCTTTCGG-----GAATACAGAGACAC 540  
 QY 206 SerLeuIrpThr-----GluSerTyrGlyGlyHisTyrGlyProAlaPheAsnHis 223  
 Db 541 AAATCTTTCGACAGGAGAGAGTATGCTGGCATCTACATCCCCAC----- 588  
 QY 224 PheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPhe 243  
 Db 589 -----TTGGCTGTACTGCTCAAGCAGGATCCTAGCATGAATCTT 627  
 QY 244 AsnSerLeuGlyIleIleAsnGlyIleIle-----AspGluAlaIle 257  
 Db 628 CAGGGCTGGCTGGGCAATGAGTTCCTCCTATGAGCAGACAGCAACTCCCTGGTC 687  
 QY 258 GluAlaProTyrTyrProGluPheAlaValAsnThrTyrGlyIleLeuAlaValAsn 277  
 Db 688 TACTTTCCTACTACCATGGCCCTTCTGGGGAACAGACTTTGGACTTCACTGCAGACCCAC 747  
 QY 278 GluThrValTyrAsnTyrMetLysPhe---AlaAsnGlnMetProAsnGlyCysGluAsp 296  
 Db 748 TGCTGGCTCAGACACAGTAGTAACTTTATGACAAACAGACCCAGAGTGTAAACAAT 807  
 QY 297 LeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAla 316  
 Db 808 CTCCTGGAAGTGTCTGAATGTGGCAAAATCTGGCTCAACATCTACATCTCTATGCT 867  
 QY 317 GluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArg 336  
 Db 868 CCG-----TGTGCTGGTGGGTGCCCGC----- 891  
 QY 337 GlyValTyrAspIleArgHisProTyrAspAsp----- 347  
 Db 892 -----ACACATAGATATGAGGACACACTGTAGTCCAGGATTTTGGCAAC 936  
 QY 347 ----- 347  
 Db 937 ATCTTCACTCGCTGCCACTTAAGCGGAGATTTCCIGAGGCACATGCTCGTCTGGGAC 996  
 QY 348 -----ProThrProProSerTyrTyr-----Asn 355  
 Db 997 AAGGTACGCTTGATCTCCTGCAACCAACACAGACCCCTTCCAACTACCTCAACAC 1056

QY 356 LysPheLeuAlaLys-----AspSerVal-----MetAspAla 366  
 Db 1057 CCTATGTTCCGAAGAGGCTCTCCACATCCCGAGTCCGCTGGGCACATGTGCAAC 1116  
 QY 367 IleGlyValAsnIleAsnTyrThr-----GlnSerAsnAspValTyrTyrAla 383  
 Db 1117 TCTGTGGTGAATTTACAGTACCGCGCTACCAAGAGCATGATCCAGTAC----- 1170  
 QY 384 PheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIleLeu 403  
 Db 1171 -----CTGAAGCTGCTCAGI 1185  
 QY 404 AlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPhe 423  
 Db 1186 TCACAGAAATCCAGATCTCTTACACCGAGATGGACATGCGCTGCAACTCATG 1245  
 QY 424 GlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAla 443  
 Db 1246 GCGCATGAGTGTGTGTGATTCGTCACACAGAGATGGAGTGCAGCGCG----- 1299  
 QY 444 GlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGlu-----Thr 458  
 Db 1300 -----CCGIGCTAGTGGAGTACGGGAGAGCGGAGACAGTACGTGGTTCTG 1350  
 QY 459 ArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyr 478  
 Db 1351 AAGGAGTGTTCACATCATCCTCTCACCATCAAGGTGCGGAGACATGTCGCCACG 1410  
 QY 479 TyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 490  
 Db 1411 GACAAGCTCGAGTGTCTTTACCAATGTTCTCGAG 1446  
 RESULT 11  
 CNS06Z5E 963 bp DNA linear GSS 07-JUL-2001  
 LOCUS T3 end of clone AY0AA015G04 of library AY0AA from strain CBS 6340  
 DEFINITION of Kluyveromyces thermotolerans, genomic survey sequence.  
 ACCESSION AL421896  
 VERSION AL421896.1 GI:12205091  
 KEYWORDS GSS.  
 SOURCE Kluyveromyces thermotolerans  
 ORGANISM Kluyveromyces thermotolerans  
 Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 REFERENCE 1 (bases 1 to 963)  
 AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
 Bollen-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,  
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,  
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
 Wincker,P. and Weissenbach,J.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)  
 MEDLINE 20584711  
 PUBMED 11152876  
 REFERENCE 2 (bases 1 to 963)  
 AUTHORS Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P.  
 and Dujon,B.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 10.  
 Kluyveromyces thermotolerans  
 JOURNAL FEBS Lett. 487 (1), 61-65 (2000)  
 MEDLINE 20584720  
 PUBMED 11152885  
 REFERENCE 3 (bases 1 to 963)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 COMMENT This GSS is part of a random genomic sequencing program of thirteen  
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,

Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

## FEATURES

```

source
    1. .963
        /organism="Kluyveromyces thermotolerans"
        /mol_type="genomic DNA"
        /strain="CBS 6340"
        /db_xref="taxon:4916"
        /clone="AY00A015G04"
        /clone_lib="AY00A"
        /note="end : T3"
misc_feature
    <270..>962
        /note="similar to Saccharomyces cerevisiae ORF YBR139w [
strong similarity to carboxypeptidase ]"
        /evidence="not_experimental"
BASE COUNT      242 a   231 c   217 g   272 t   1 others
ORIGIN
Alignment Scores:
Pred. NO.:      2,42e-28      Length:      963
Score:          352.50      Matches:      88
Percent Similarity: 45.08%      Conservative: 31
Best Local Similarity: 33.33%      Mismatches:  98
Query Match:     12.21%      Indels:      47
Db:              29          Gaps:         5

US-09-712-338-2_COPY_19_555 (1-537) x CNS0625E (1-963)
Qy      1  LeuProGlySerThrProAlaSerValGly-----ArgArgGlnLeuProLysAsnPro 18
Db      297  GTCCCAAAATACCGTGCAGTCTATCAAAATGCGTGAATCTGGAATCTGGACGAGTTGCTC 356
Qy      19  ThrGlyValIysThrLeuThrAlaAsnValThrIleArgTyrLysGluProGly 38
Db      357  ACGATAAGTGTGAGCTGTGATGACACCTACTCGATGAGGTAAGAGAGAGTGGACCCCTCG 416
Qy      39  AlaGluGlyValCysGluThrThrProGlyValIysSerTyrSerGlyTyrValAspThr 58
Db      417  AAGTGGGTGTC-----GATCTGTCAAGCAGTACTCGGGGTACTCTGGGACTCTGGACATAC 464
Qy      59  SerProGluSerHisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAla 78
Db      465  GAGGATTCGAACACACTCTTTTACTGGGCATTTGAATCCAGAAACGACCCACTCAATGAC 524
Qy      79  ProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGlu 98
Db      525  CCCGTTATTCTATGGCTGAAGCGTGTCCAGGGTGTCTCTTTTCAAGGGGTATTATTTC 584
Qy      99  GluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrp 118
Db      585  GAATCTGGCCCTCTCCGTCGGCCCTGAGCTAAGCCCGTTCGCAATCCGCTACTCCGTCG 644
Qy      119  AsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSer 138
Db      645  AATAACAATGCTACTGTCATATCTTGGAGCAGCCCTTAGCGGTGTGTTTTCCTATGCT 704
Qy      139  AspThrValAspGlySerIleAsnProValThrGlyValGluValGluAsnSerPheAla 158
Db      705  GAC-----
Qy      159  GlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAla 178
Db      708  -----GAGAGAGTGGCTCCACATACGACGACGAC 734
Qy      179  AlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSer 198
Db      735  GGAAGGATGTGTTTCATATCTTGGAGTGTGTTTTCAGAGATGTTCCCAAGATGTTCCCA----- 782

```

```

Qy      199  ArgValGlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyPro 218
Db      783  CAGTTTAGATCTAAGCACTTCCATATTCAGTGCAGTCAATCAAGTGCAGTCAATCAATCA 842
Qy      219  AlaPhePheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGly 238
Db      843  -----GAGATTGCCCATCAATPCGCTGTCGTTCTCAATCAATCAATCAATCAATCA 881
Qy      239  ValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAsnGluAlaIleGln 258
Db      882  AAACCTTCATCTCACTCCTCATCATGATGTCGCAACGGGATACAGATCTCTTAGTCCAG 941
Qy      259  AlaProTyrTyr 262
Db      942  TATGAMTACTAT 953

RESULT 12
AK019861
LOCUS
DEFINITION
    AK019861      2493 bp      mRNA      linear      HTC 05-DEC-2002
    Mus musculus 11 days pregnant adult female ovary and uterus cDNA,
    RIKEN full-length enriched library, clone:5031400F07
    product:protective protein for beta-galactosidase, full insert
    sequence.
ACCESSION
    AK019861
VERSION
    AK019861.1  GI:12860233
KEYWORDS
    HTC; CAP trapper.
SOURCE
    Mus musculus (house mouse)
ORGANISM
    Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
    1  Carninci,P. and Hayashizaki,Y.
        High-efficiency full-length cDNA cloning
        Meth. Enzymol. 303, 19-44 (1999)
        99279253
        10349636
REFERENCE
    2  Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
        Itoh,H., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
        Normalization and subtraction of cap-trapper-selected cDNAs to
        prepare full-length cDNA libraries for rapid discovery of new genes
        Genome Res. 10 (10), 1617-1630 (2000)
        20499374
        11042159
REFERENCE
    3  Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
        Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
        Sumi,N., Ishii,Y., Nakamura,S., Hazawa,M., Nishine,T., Harada,A.,
        Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
        Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M.,
        Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
        Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
        RIKEN integrated sequence analysis (RISA) system--384-format
        sequencing pipeline with 384 multicapillary sequencer
        Genome Res. 10 (11), 1757-1771 (2000)
        20530913
        11076861
REFERENCE
    4  Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
        Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
        Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
        Saito,T., Okazaki,Y., Gojohori,T., Bono,H., Kasukawa,T., Saito,R.,
        Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
        Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
        Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
        Quackenbush,J., Schriml,L.M., Stauble,F., Suzuki,R., Tomita,M.,
        Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
        Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
        Carninci,P., de Bonaldo,M.P., Brownstein,M.J., Bult,C.,
        Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
        Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
        Marchionni,L., Mashima,J., Mazzarelli,J., Monbaerts,P., Nordone,P.,
        Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,

```

```

CDS
284..1708
/dev_stages="11 days pregnant adult"
/note="unnamed protein product; protective protein for
beta-galactosidase (MGD|MGI:97748)
putative"
/codon_start=1
/protein_id="BAB31888.1"
/db_xref="GI:12860234"

```

Db 980 ACCACATGTCGCGCTCAGACAAAGTGTAACTTATGACACAAAGACCCAGAGTGTA 1039  
 Qy 295 GlnAspLeuIleSerThrCysLysGluThrAsnArgThrAlaLeuAlaAspTyrAlaLeu 314  
 Db 1040 AACAAATCCUGGAAGTGTCTGAAATGTGGCAATCTGGCCCTCAACATCAATCAATCTC 1099  
 Qy 315 CysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAla 334  
 Db 1100 TATGCTCCG-----TGTCGTGGTGGGTGCCGCGC----- 1129  
 Qy 335 GlyArgGlyValTyrAspIleArgHisProTyrAspAsp----- 347  
 Db 1130 -----AGCATAGATATAGGACACACTGTAGTCCAGCAATT 1168  
 Qy 347 ----- 347  
 Db 1169 GGCAACATCTTCACTCGCTGCCACTTAAGCGGAGATTTCTCTGAGGCACATGCGTTCT 1228  
 Qy 348 -----ProThrProProSerTyrTyr--- 354  
 Db 1229 GGGGACAAGGTACGCTTGGATCCTCCCTGCAACCAACACACAGCCCTTCCAACTACCTC 1288  
 Qy 355 ---AsnLysPheLeuAlaLys-----AspSerVal-----Met 364  
 Db 1289 AACAAACCCCTATGTCGAGAGCTCTCCACATCCCGAGTGGCTGCCCCCGCTGGGACATG 1348  
 Qy 365 AspAlaIleGlyValAsnIleAsnTyrThr-----GlnSerAsnAspValTyr 381  
 Db 1349 TGCAACTTCTTGGTGAATTTACAGTACCGCGCTCTACCAACCATGATCCAGCTAC 1408  
 Qy 382 TyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu 401  
 Db 1409 -----CTGAAGCTG 1417  
 Qy 402 IleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsn 421  
 Db 1418 CTCATGTCAGAAATACCAATCTCTGCTACAAAGAGATGTCGACATGCGCTGCAAC 1477  
 Qy 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGlnPheArg 441  
 Db 1478 TTCATGGCGCATGAGTGGTTGTGGATTCGCTCAACCAAGAGATGGAGGTGCAGCGCGG 1537  
 Qy 442 SerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGlu----- 457  
 Db 1538 -----CCCTGGCTAGTGGCTACGGGAGAGCGGAGCAACAGGTACGCTGG 1582  
 Qy 458 ---ThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluVal 476  
 Db 1583 TTCGTGAAGGAGTGTTCACATCACTCTCTCACCATCAAGGTGCGCGACATGCTC 1642  
 Qy 477 ProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 490  
 Db 1643 CCCACGGACAAGCTCGAGCTGCTTTTACCATGTTCTCGAGG 1684

## RESULT 13

CA747625  
 LOCUS GAP13D07 Aspergillus flavus Lambda zap cDNA Library Aspergillus  
 DEFINITION flavus cDNA, mRNA sequence.  
 ACCESSION CA747625  
 VERSION 1  
 KEYWORDS EST.  
 SOURCE Aspergillus flavus  
 ORGANISM Aspergillus flavus  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 1 (bases 1 to 721)  
 AUTHOR Orian, G.R., Fakhoury A.M. and Payne, G.A.  
 TITLE Identification of Genes Differentially Expressed During Aflatoxin  
 Biosynthesis in Aspergillus flavus and Aspergillus parasiticus  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Payne, GA  
 Department of Plant Pathology

North Carolina State University  
 Box 7616, Gardner Hall, Raleigh, NC 27695, USA  
 Tel: (919) 515-6994  
 Fax: (919) 515-7716  
 Email: Gary\_Payne@NCU.

## FEATURES

source  
 1..721  
 /location/Qualifiers  
 /organism="Aspergillus flavus"  
 /mol\_type="mRNA"  
 /strain="3357"  
 /db\_xref="taxon:5059"  
 /clone\_lib="Aspergillus flavus Lambda Zap cDNA Library"  
 /note="vector: pbluescript; Log phase of aflatoxin  
 production"

BASE COUNT 161 a 159 c 224 g 176 t 1 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 1,11e-27 Length: 721  
 Score: 345.00 Matches: 66  
 Percent Similarity: 61.54% Conservative: 30  
 Best Local Similarity: 42.31% Mismatches: 60  
 Query Match: 11.95% Indels: 0  
 Db: 14 Gaps: 0

US-09-712-338-2\_COPY\_19\_555 (1-537) x CA747625 (1-721)

Qy 369 ValAsnIleAsnTyrThrGlnSerAsnAspValTyrTyrAlaPheGlnGlnThrGly 388  
 Db 8 GTCCCTGCAACTTCCACGGAAGCTGCCCAAGTGTCCCAACAAAGATTTCACAGACCTTAC 67  
 Qy 389 AspPheValTrpProAsnPheIleGluAspLeuGluGlnIleLeuAlaLeuProValArg 408  
 Db 68 GACATCACCCGCGCGGATTCCTCGACTCGATAGCCACCTCTCGACAGCGGCGTCAAG 127  
 Qy 409 ValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTyrPheGlyGlyGlnAlaVal 428  
 Db 128 GTACACATGATGTACGGCGACCGGACTACCGCTGTAACTGGATCGCGGCGGAGAAAGCC 187  
 Qy 429 SerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeu 448  
 Db 188 AGCTTCGACGTACCGTACTCGCGGCGCGCAGATTCGCAACACCGGCTNACTCCATTA 247  
 Qy 449 LysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArg 468  
 Db 248 GTGACGTCGGAAGCATCAAGGTATGACAGCTCAACTCACTTAATATAGCTTCACGCGC 307  
 Qy 469 ValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPhe 488  
 Db 308 GTCACCAGGCTGTGTCATGAAGTCCGCTTATCAGCGGCTTATCAGCGGCTATGAGATCTTC 367  
 Qy 489 AsnArgThrIlePheGlyTrpAspIleAlaGluGlnLysLysIleTrpProSerTyr 508  
 Db 368 ATGGTGTCTAGCTGGATCGGATATTCCTACTGGAGAGATAGGTATTTACCGGGGAGTTT 427  
 Qy 509 LysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSerValPro 524  
 Db 428 AAGACGAGTGACCGGAGGATACCTTGGCATATTAAAGATTTGTGCTCCA 475

## RESULT 14

AF217508  
 LOCUS AF217508  
 DEFINITION Homo sapiens uncharacterized bone marrow protein BM031 mRNA,  
 complete cds.  
 ACCESSION AF217508  
 VERSION AF217508.1 GI:7688958  
 KEYWORDS HTC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1581)  
 REFERENCE Zhang, Q.H., Ye, M., Wu, X.Y., Ren, S.X., Zhao, M., Zhao, C.J., Fu, G.,  
 AUTHORS



Shen, Y., Fan, H. Y., Lu, G., Zhong, M., Xu, X. R., Han, Z. G., Zhang, J. W.,  
Tao, J., Huang, Q. H., Zhou, J., Hu, G. X., Gu, J., Chen, S. J., and Chen, Z.  
Cloning and functional analysis of cDNAs with open reading frames  
for 300 previously undefined genes expressed in CD34- hematopoietic  
stem/progenitor cells

Genome Res. 10 (10), 1546-1560 (2000)

JOURNAL  
MEDLINE  
PUBMED

20499367  
11042152

REFERENCE

2 (bases 1 to 1581)

Zhao, M., Gu, J., Li, N., Peng, Y., Han, Z., and Chen, Z.

A novel gene expressed in human bone marrow

Unpublished

3 (bases 1 to 1581)

Zhao, M., Gu, J., Li, N., Peng, Y., Han, Z., and Chen, Z.

Direct Submission

Submitted (20-DEC-1999) Chinese National Human Genome Center at

Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,

Shanghai 201203, P. R. China

Location/Qualifiers

FEATURES

source

1..1581

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/tissue\_type="bone marrow"

355..1056

/codon\_start=1

/product="uncharacterized bone marrow protein BM031"

/protein\_id="AA67619.1"

/db\_xref="GI:7688959"

/translation="MFGIFVEHGVVVTNNMLRDPPFWTTLSMLYIDNPVGTGFS

FTDTHGYAVNEDVDALYSALIOFFQIPPEYKNDYFVTGSEYAKYVPAIAHLIH

SLNPVRVKNLNGIALGDGYSOPESLIGGYAEFLYQIGLIDEXOKYFKQCHECIE

HIRKQNFEEFELDKLIDGDISDPYSFQNVTCSSNYNFLRCIEPEDQLYYKFLS

LPEVQAIPKGSDF"

BASE COUNT 473 a 317 c 362 g 429 t

ORIGIN

Alignment Scores:

Pred. No.: 7,11e-26 Length: 1581

Score: 333.00 Matches: 130

Percent Similarity: 39.34% Conservative: 62

Best Local Similarity: 26.64% Mismatches: 178

Query Match: 11.53% Indels: 118

DB: 11 Gaps: 21

US-09-712-338-2\_COPY\_19\_555 (1-537) x AF217508 (1-1581)

Qy 47 Progly-----ValLysSerTyrSerGlyTyr-----ValAspIhrSerProGluSer 62  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
Db 204 CCAGGACTGAACATGAAGAGTATGCGCGCTTCCTCACCGGTGAATAAGACTTACACAGC 263  
Qy 63 HisThrPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIle-ThrLe 82  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
Db 264 AACCT 323  
Qy 82 utrpLeuAsnGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPr 102  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
Db 324 CTGCTACAGGTGGCGGGAGGTTCATCCATGTTGGACTTTTGTGGAACATGGGCC 383  
Qy 102 oCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluVal 122  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
Db 384 TTATGTTCTCAAGTAACATGACCTTGGGTGACAGAGACTTCCCTGGACCACACAGCT 443  
Qy 122 rAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerThrValAs 142  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
Db 444 CTCATGCTTTACATTCACATTCACATTCACATTCACATTCACATTCACATTCACATTC 503  
Qy 142 pGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGln 162  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
Db 504 CGGA----- 507  
Qy 162 yArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAla 182  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
Db 1404 TTTAIGGAAAGGATGGAT 1423

Db 508 -----TATGCAGTCAATGAGCAGCATGTAGCAGCGGATTATATA 545  
Qy 182 atrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSe 202  
Db 546 CAGTGCACATATTCAGTTTTC-----CAGATATATTCCTGAATATAFAAAA 590  
Qy 202 rLysAspPheSerLeuTrpThrGluSerIleGlyHisTyrGlyProAlaPheAs 222  
Db 591 TAATGACATTTATGTCAGTGGGAGTCTATGCGGGGAATATGCGCAGCATATGCACA 650  
Qy 222 nHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAs 242  
Db 651 CCTCATCATTCCTCAAC-----CCTGTGAGAGAGGTGAAGATCAA 692  
Qy 242 nPheAsnSerLeuGlyIleIleAsnGlyIleLeuAspGluAlaIleGlnAlaProTyrTy 262  
Db 693 CTGAACAGGAATTCGATATGATGATGATATCTGATCCGGAATATATATAGGGGGGTA 752  
Qy 262 rProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyAs 282  
Db 753 TGCAGAAATC-----CTGTACCAATTTGGCTTTGGGATGAGAGCAAGCAAAA 800  
Qy 282 nTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLy 302  
Db 801 GTAC-----TTCCAGAAAGCAG-----TGCCATGAATGCATAGACACAICAG 842  
Qy 302 sGlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaTh 319  
Db 843 GAAGCAGAACTGGTTTGGAGGCTTTGAATACTGATAAATACTAGTAGTGGCAGCTTAAC 902  
Qy 319 rAsn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaG 335  
Db 903 AAGGTATCTCTTACTTCCAGATGTACAGAGTGTAGTAACTACTACTAATCTTTTG-- 960  
Qy 335 yArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyAs 355  
Db 961 -----CGGTGCAGCGAACCTGAGGATCAGCTTACIATGT 995  
Qy 355 nLysPheLeuAlaLysAspSerValMetAspAlaIle-GlyValAsnIleAsnTyrThrG 375  
Db 996 GAAATTTTGTCACTCCAGAGGTGAGACAGCCATCCAGCTGGGG---AATCAGACTT 1052  
Qy 375 LnSerAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValIrrp 393  
Db 1053 TTAATGATGGAACTATAGTTGAAAGTACTTGGGAGAAGATACATACAGTCAGTTAAGC 1112  
Qy 393 roAsnPheIleGluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIle 413  
Db 1113 CATGTTTAACTGAATCATGAATAAT-----TATAAGCTTCTGATCTACA 1157  
Qy 413 yrGlyAspAlaAspTyrIle-----CysA 421  
Db 1158 ATGCCAACTGGACATCACTGCGCAGCTGCCCTGACAGAGCGCTCTTGTGGSCATGG 1217  
Qy 421 snTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaGln---- 439  
Db 1218 ACTGGAAGGATCCAG-----GAATACAGAGAGGAGCAAGAAAAG 1259  
Qy 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnG 452  
Db 1260 TTTGGAAGATCTTTAAATCTGACAGTGAAGTGGCTGGTTAC----- 1300  
Qy 452 lyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyGluA 472  
Db 1301 -----ATCGGGAAGTGGGTGACTTCCATCAGGTAATTAATTCAGGTG 1343  
Qy 472 laGlyHisGluValProTyrThrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr 492  
Db 1344 GAGGACATATTTACCCCTATGACCGCTCTGAGAGCTTGTGACAIGATTAATGATTC 1403  
Qy 492 lePhe-----GlyTrpAsp 496  
Db 1404 TTTAIGGAAAGGATGGAT 1423

RESULT 15  
 LOCUS CD045123 712 bp mRNA linear EST 09-MAY-2003  
 DEFINITION psHB017xN05f.183846 psHB: Infected hypocotyl soybean host. 48 hrs post infection Phytophthora sojae cDNA clone SHB017N05 5', mRNA sequence.

ACCESSION CD045123  
 VERSION CD045123.1 GI:30498716  
 KEYWORDS EST.  
 SOURCE Phytophthora sojae  
 ORGANISM Phytophthora sojae  
 Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora.

REFERENCE 1 (bases 1 to 712)  
 AUTHORS Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.  
 TITLE USDA-IFAFS: Expression of Phytophthora sojae genes during infection and propagation  
 JOURNAL Unpublished  
 COMMENT Contact: Tyler B  
 Tyler lab  
 VBI  
 1880 Pratt Dr., Blacksburg, VA 24061, USA  
 Tel: 540-231-7318  
 Email: bmtyley@vt.edu  
 PCR Primers  
 FORWARD: BK reverse  
 Plate: 017 row: N column: 05  
 Seq primer: BK reverse  
 High quality sequence stop: 712.  
 Location/Qualifiers  
 1..712  
 /organism="Phytophthora sojae"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:67593"  
 /clone="SHB017N05"  
 /tissue\_type="infected host tissue"  
 /cell\_line="P6497"  
 /dev\_stage="48 hour post infection"  
 /clone\_lib="psHB: Infected hypocotyl soybean host. 48 hrs post infection"  
 /note="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI; USDA-IFAFS: Expression of Phytophthora sojae genes during infection and propagation."

BASE COUNT 154 a 224 c 218 g 116 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2.7e-26 Length: 712  
 Score: 332.50 Matches: 82  
 Percent Similarity: 44.65% Conservative: 39  
 Best Local Similarity: 30.26% Mismatches: 97  
 Query Match: 11.52% Indels: 53  
 DB: 14 Gaps: 8

US-09-712-338-2\_COPY19\_555 (1-537) x CD045123 (1-712)

QY	10	GlyArgGlnLeuProLysAsnProThrGlyValLysThrLeuThrThrAlaAsnAsn	29
Db	42	GGCTGGCGGCCATCGAAGACAGTGGCGGTGGTGCCTACCGTCACTGCTGGGGCGAC	101
QY	30	ValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyVal	49
Db	102	ACAGTC-----ATGTGTGGCACC-----GCA	122
QY	50	LysSerTyrSerGlyTyrValAspThr-----SerProGluSerHisThrPhePheIrp	67
Db	123	ACCAACGAAGCGGCTACGTCAAGTGGCCGACACAGCAGCAGCATCTCTACTGG	182
QY	68	PhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyGly	87
Db	183	TTCTTCAGTCCGCGAGAGCTCTGCTACCGACCCCTGCTGCTGCTTGAAGCGGTGC	242

Search completed: September 17, 2003, 01:44:29  
Job time : 3392 secs

88	ProCysSerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsnSer	107
243	CCCGAGATTCCACGCCCTATGACGCTCTIGACCAGACGCGCGTGCTGTTAAAGAG	302
108	ThrPheAspAspTyrIleAsnProHisSerThrAsnGluValSerAsnLeuLeuPheLeu	127
303	GATATGTCGAGGAGGCCAACCACACCTCTGGAACTCGGAGGCCAAGTCATCGGCTG	362
128	SerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnPro	147
363	GACCAGCCCGACGACGTCGGTACTCGTACGCGACCCCGCCGAGTCCGCAC	419
148	ValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIle	167
420	GAGAAAGACGTCACAGAGAAC-----	440
168	AspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTyrGluIleLeuGln	187
441	-----GTCACGGCTTCCTACAG	458
188	GlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPheSerLeu	207
459	GGGTTCGGACACACCCCGAGCTT-----CAAGACGGACCGCTGTCCTC	506
208	TyrThrGluSerTyrGlyHisTyrGlyProAlaPheAsnHisPheTyrGluGln	227
507	CGAGCGAGAGCTACGCGAGGCCACTATACCCAGCAGCGCGGCACAGATCCACCGCAG	566
228	AsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeuGly	247
567	AACAAA-----GCGCGAAGTCAGGCGCTCAAACTCCAAAGGCATGCC	611
248	IleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAlaVal	267
612	ATCGGGAACGGCTTGACCAACACCGCTCGTGACAGCGGACGACGGCTGGCATG---	668
268	AsnAsnThrTyrGlyIleLysAlaValAsnGlu	278
669	AACAACTCGTACGGCGTCAAGCTCATGGACAC	701